Tabular Information Extraction From Datasheets With Deep Learning for Semantic Modeling

by

Yakup Akkaya

Thesis submitted to the University of Ottawa
in partial fulfillment of the requirements for the
degree of Master of Applied Science

School of Electrical Engineering and Computer Science
Faculty of Engineering
University of Ottawa

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

The growing popularity of artificial intelligence and machine learning has led to the adoption of the automation vision in the industry by many other institutions and organizations. Many corporations have made it their primary objective to make the delivery of goods and services and manufacturing in a more efficient way with minimal human intervention. Automated document processing and analysis is also a critical component of this cycle for many organizations that contribute to the supply chain. The massive volume and diversity of data created in this rapidly evolving environment make this a highly desired step. Despite this diversity, important information in the documents is provided in the tables. As a result, extracting tabular data is a crucial aspect of document processing.

This thesis applies deep learning methodologies to detect table structure elements for the extraction of data and preparation for semantic modelling. In order to find optimal structure definition, we analyzed the performance of deep learning models in different formats such as row/column and cell. The combined row and column detection models perform poorly compared to other models' detection performance due to the highly overlapping nature of rows and columns. Separate row and column detection models seem to achieve the best average F1-score with 78.5% and 79.1%, respectively. However, determining cell elements from the row and column detections for semantic modelling is a complicated task due to spanning rows and columns. Considering these facts, a new method is proposed to set the ground-truth information called a content-focused annotation to define table elements better. Our content-focused method is competent in handling ambiguities caused by huge white spaces and lack of boundary lines in table structures; hence, it provides higher accuracy.

Prior works have addressed the table analysis problem under table detection and table structure detection tasks. However, the impact of dataset structures on table structure detection has not been investigated. We provide a comparison of table structure detection performance with cropped and uncropped datasets. The cropped set consists of only table images that are cropped from documents assuming tables are detected perfectly. The uncropped set consists of regular document images. Experiments show that deep learning models can improve the detection performance by up to 9% in average precision and average recall on the cropped versions. Furthermore, the impact of cropped images is negligible under the Intersection over Union (IoU) values of 50%-70% when compared to the uncropped versions. However, beyond 70% IoU thresholds, cropped datasets provide significantly higher detection performance.

Acknowledgements

I would like to express my deepest gratitude to my supervisor Dr. Burak Kantarci for his guidance and endless support. From the first time we met, he has always been patient and polite in a solution-oriented manner regarding research issues and any other problems. Regardless of the time, it is remarkable that he can be reached for a discussion online or face-to-face over an e-mail. I am grateful for the opportunity to work with him as well as for everything he has done for me.

I also want to thank Dr. Murat Simsek for motivating me to pursue a master's degree in a different field and introducing me to Dr. Burak Kantarci. His constructive criticism, ideas, and assistance play an important role in my work. I would also like to thank Johan Fernandes, JiChu Jiang and Bin Xiao for their help and efforts. Their insight and perspective were vital in my achievement.

Lastly, I would like to express my gratitude to my dear wife, Zeyneb Akkaya, who has been supportive and understanding throughout my working day and night. Despite our great distance, I would like to express my profound gratitude to my family for their unwavering support. They have stood by my shoulder through all of my challenges and have been my strongest supporters in every decision I have made. Their absolute belief in me has been the most important aspect in getting me to where I am now. Finally, I owe my deepest gratitude to my dear brother Gokhan Sancak. He was instrumental in my decision to come to Canada, a watershed moment in my life, and has always been my greatest motivator.

Table of Contents

List of Tables

List of Figures

Publications of the Candidate During MASc Studies

Outcomes of the thesis:

- **Y. Akkaya**, M. Simsek, B. Kantarci, and S. Khan "On Cropped versus Uncropped Training Sets in Tabular Structure Detection," *Elsevier Neurocomputing* (Under Review)

Chapter 1

Introduction

The amount of unstructured data pushed by billions of connected devices has become overwhelmingly high in a variety of categories and forms with the advent of Industry 4.0 paradigm. Alongside, artificial intelligence, machine-to-machine communication, and Internet of Things (IoT) devices are being incorporated into automated manufacturing to minimize human intervention in production [1]. Automation of data processing and analysis is also a critical component of this cycle for many organizations that contribute to the supply chain [2]. Tables in documents with huge volumes of data often contain valuable information such as cost, technical specifications, requirements, and performance capabilities. Automated extraction of this information from documents is of paramount importance particularly to pave the way for optimized supply chain management. The state of the art offers image recognition-based approaches with Convolutional Neural Network (CNN) classifiers as a solution to this problem. However, heterogeneity of table structures and layouts remain a challenge for detection.

Extracting tabular information out of the image documents involve two separate tasks. The first step is to detect the table with the position on the document. Once the table region is captured, the structure of the table is detected. This task can be executed by segmenting the table layout as rows and columns or simply as table cells. Prior to the deep learning-based solutions, existing methods were largely operating on on digital-born PDF documents (which retained all PDF meta-data). PDF-based approaches exploit the meta-data associated with documents in combination with heuristic rules [3, 4]. However, PDF documents may not contain metadata with standardized structure since most of the documents are created from scanned images.

Deep learning based approaches have been popular for table detection and structure

detection. Works that have been published on structure detection are relatively much rarer than those on table detection [5, 6, 7, 8, 9, 10, 11], due to the more complicated nature of the problem. The lack of standards for tables, textual regions and other figures within the documents make determining table objects challenging. Recognizing the anatomy of the table is a more challenging task than table detection as it requires detecting irregularities such as spanning rows and columns, and the identification of merged cell objects in complex table structures. Rows and cells can be quite densely laid out in a page, and have tiny dimensions. Besides, both tasks require high volumes of hand labelled data, and hand-labelling is a time-consuming task, and thus, these useful training resources are scarce.

Many studies have adopted Region-Based Convolutional Neural Network (R-CNN) algorithms in table detection and structure detection. The presented works in [12, 13] achieved satisfactory results with over 96% F1-score for row and column detection by using Mask R-CNN. Prasad et al. [14] utilized Cascade Mask R-CNN, which is developed upon Mask R-CNN and achieved the highest accuracy results on the ICDAR 2019 structure detection dataset. Jiang et al. [15] show that Hybrid Task Cascade with dual ResNeXt101 backbone outperforms the existing solutions by providing around 8%-9% higher F1-Score (81.8%) than Mask R-CNN and Cascade Mask R-CNN in table structure detection task in cell recognition.

1.1 Motivation

The tables in the documents contain valuable information for a variety of fields. In document processing, obtaining this information in an automated manner and eliminating human intervention is highly desired. In deep learning-based solutions, the problem is addressed under two separate tasks. Tables and their structure in document images are accepted as objects and aimed to be localized by object detection algorithms. Table detection and table structure detection tasks form the process of extracting tabular information from documents in the literature. Existing studies either present an end-to-end approach to perform both tasks or directly focus on structure detection. Most structure detection related studies assume that tables are perfectly localized and focused on the detected table area. However, the impact of having a table detection model on the performance of structure recognition task has not been discussed in the literature. In other words, a model can seek table structure in the whole document image instead of the detected table area. Thus deep learning model might generalize better on table structure by learning with surrounding textual regions and other figures.Consequently, a robust structure detection model can make the table detection step impractical. On the other hand, regardless of the

robustness of a structure detection model, the table detection step may provide additional improvement by restricting the search area to the detected region. Therefore a comparison between the two approaches is beneficial.

Another critical part of the problem is the definition of table structure. There is no precise and globally defined standard format for the structure of tables. They can appear in a variety of styles and have a complex structure that makes the definition of table structure challenging. Hence, analysis of table structure definitions in various formats, e.g., row, column and cell, is essential to find the optimal solution. Since deep learning models are data-driven, error-free and consistently annotated data is a great need for table structure detection. The ground-truthing table structure information is tedious, time-consuming, expensive as well complicated due to ambiguities. The quality of labelled data and the way it is defined is a key factor for the performance of deep learning models. The complexity of the problem can be reduced to the minimum, and an efficient strategy can provide a fast and accurate labelling process.

1.2 Objectives

In this thesis, we assume that the tables are perfectly localized and focus on detecting structure. The primary goal is to segment table structure into its fundamental components for automated tabular data extraction from documents. To achieve this, we investigate the optimal format to define the table structure most simply. Different structure formats such as rows and columns, and cells are studied and evaluated with numerical results as well as visual outputs. Then we analyze the most efficient training scheme for structure detection models with the cropped and uncropped approaches. An efficient annotation bootstrapping strategy is introduced to overcome the challenges of labelling table structure, and problems are discussed in detail. We propose a content-focused cell detection model to accurately detect table structure and create the most straightforward output for semantic modelling.

1.3 Contributions

Following a thorough review of the literature, experiments were conducted to resolve uncertainties in the definition table structure and find the best scheme for the training of deep learning models. We propose content-focused cell detection by using cropped training sets along with an annotation bootstrapping strategy.

Our main contributions can be summarized as follows;

- We propose a content-focused cell detection method that resolves the issues of other investigated methods. The proposed method reduces the performance degradation effect of the partial detection with large white spaces due to the nature of the IoU metric. The row and column detection models do not handle multiple row/column contents adequately, and it requires two different specialized models since row and column objects are highly overlapping. Also, this thesis proposes an efficient annotation bootstrapping strategy to accelerate the labelling process and reduce erroneous ground truth to a minimum. We utilize the existing PDF-based table detection tool Pdfplumber [16] to obtain content-focused cell bounding boxes and map them into the desired format according to the corresponding image document. ICDAR 2013 and private Lytica Inc. dataset that contains 1000 document images are annotated in content-focused cell format with the proposed method.

- We propose cropped training schemes instead of uncropped scheme for deep learning models. The cropped training scheme can improve detection performance by up to 9% in average precision and recall. The impact of the cropped scheme is negligible under the IoU thresholds of 50%-70% compared to the uncropped scheme. On the other hand, Cropped datasets yield significantly better detection performance above 70% IoU values.

1.4 Thesis Outline

This thesis is comprised of six chapters. In Chapter 2, we explore the literature for studies that expressly or indirectly address the table structure recognition problem. The reviewed studies are grouped under two different titles. In Section 2.1, solutions that have traditional approaches for table structure recognition are presented in detail. Traditional approaches often combine visual indicators with rule- and heuristic-based methods to recognize table elements. Also, machine learning-based solutions are considered traditional methods. They are usually used to classify extracted features from documents on top of the hand-crafted rules. In Section 2.2, deep learning approaches based on CNNs and variants have been presented for table structure detection and recognition.

Chapter 3 delves into deep learning methodologies that we adapted in our work. We start by introducing the region-based convolutional neural networks that shape the two-staged object detection algorithms. Initially, we go through Fast R-CNN [17] and Faster R-CNN [18] in depth because they are the foundation for object detection techniques that are employed in this work. Then, we introduce Mask R-CNN [19] that is built on Fast R-CNN

where a mask branch is added in parallel to bound box regressor and softmax classifier with a few modifications. Following, Cascade Mask R-CNN [20], which is built by sequential design of multiple detection heads, and Hybrid Task Cascade (HTC) [21] structure that improves Cascade Mask R-CNN by exploiting the relations between detection head and segmentation head are reviewed. Two-staged object detectors use backbone networks for feature extraction from input images. Lastly, an overview of these backbone networks is provided.

In Chapter 4, we provide a detailed overview of table structure definition and explain the challenges in ground-truthing. First, different formats are presented to segment the table into its core components. We explain the detection of table elements in row/column and cell formats and how ground-truth information is determined. Then, training and dataset details of experiments are given to find the optimal structure format. Following, numerical and inference results of different table structure formats are evaluated. Finally, we provide the pipeline of the proposed efficient annotation bootstrapping strategy for content-focused cell labelling.

In Chapter 5, we analyze different training schemes for table structure detection. First, the cropped and uncropped approaches for structure detection models are explained. Then, the training and dataset details for experiments are given. We present the findings of two sets of experiments for cropped and uncropped training schemes with numerical and visual data. Lastly, we summarize the analysis of cropped and uncropped training schemes.

Finally, Chapter 6 wraps up the approaches described in the thesis, bestows a conclusion.

Chapter 2

Literature Survey and Background

Nowadays, documents exist in various formats and are created in massive amounts across a wide range of fields. These documents can be generated as digital-born PDFs or scanned images. The critical information within the diverse documents is mainly presented in tables since they cannot be put into words by sentences in a structured way with relevant key-value pairs. Hence, extracting the information from tables is in high demand and an essential task for document processing. To achieve the information retrieval process in an automated manner, table detection and table structure recognition tasks have been studied in document analysis and understanding for a long time.

Many methods, i.e. rule-based, machine learning and deep learning, have been used to extract tabular information. Those methods are grouped under two categories, namely traditional approaches and deep learning approaches. Pipeline of traditional and deep learning approaches are presented in Fig.2.1. In both approaches, table content is extracted in two stages: 1) table detection, 2) table structure recognition. Deep learning solutions mainly address the problem sequentially in the mentioned order. It is called as top-down approach since the detected table image decomposed into fundamental objects such as row/column or cell. However, some works propose to identify line separators or table components for the detection of the table. It is called as bottom-up approach, where tables are built from their fundamental objects.

Although far more research studies focus on table detection alone in the past, attempts to describe work done to perform table structure recognition have an increasing trend [22] in recent years. This chapter will focus on works that perform structure recognition with traditional and deep learning approaches.

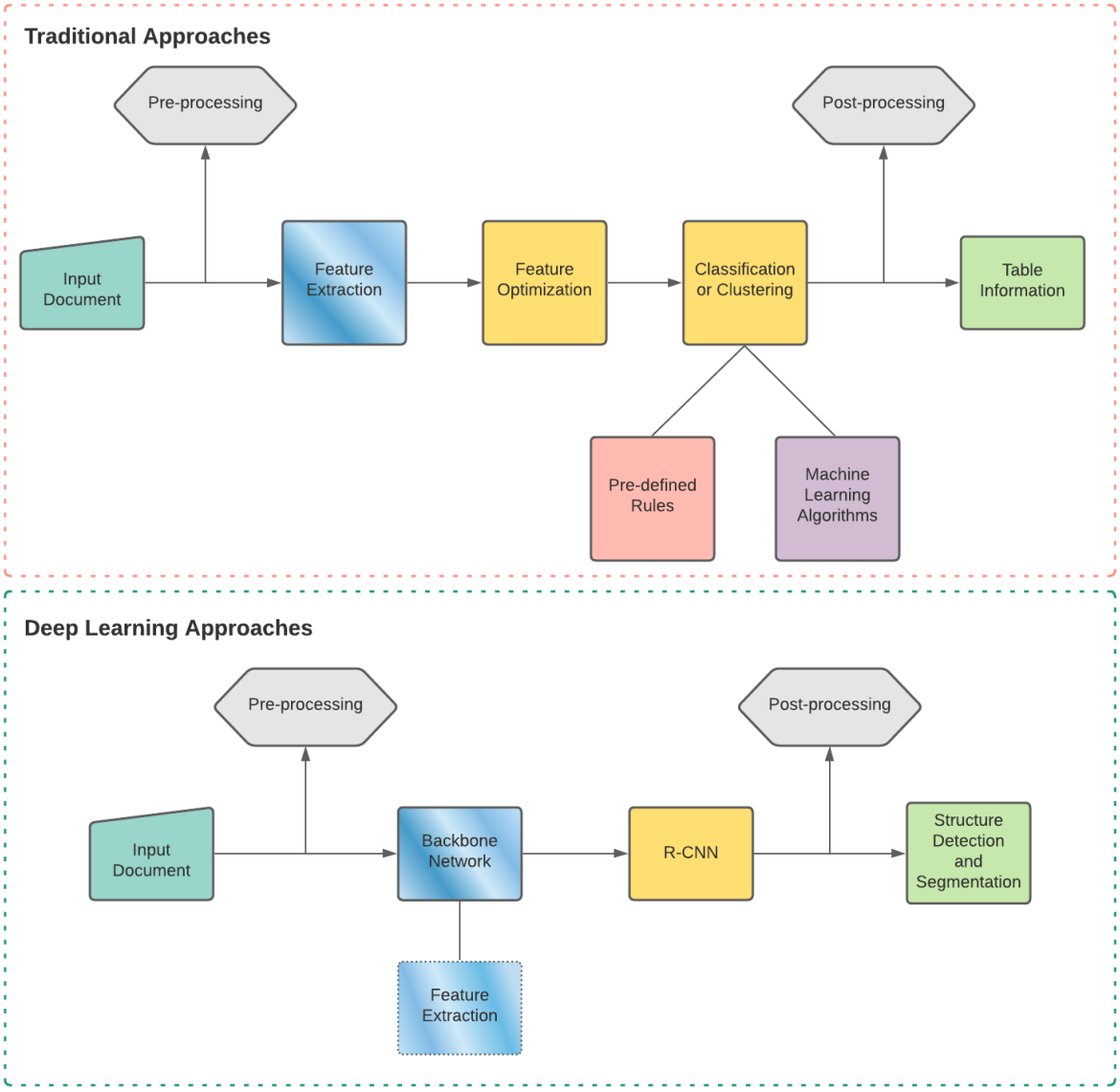


Figure 2.1: Pipeline of Traditional and Deep Learning Approaches

2.1 Traditional Approaches

Proposed solutions for table structure detection usually either work on document images or PDFs. Traditional approaches on document images mainly utilize visual indicators such as vertical and horizontal separator lines and blank spaces between table components. These visual clues are used by the rule-based solutions and pawed the way for more advanced systems. Another method is the vertical segmentation of word blocks along with neighbour word relations, alignments, and intercolumn spaces. On the other hand, PDF-based methods heavily rely on metadata associated with documents. The features that are used in rule-based solutions are obtained from this metadata. In addition to these, machine learning solutions have proven to be a strong candidate, to replace hand-crafted rules or push it to be part of post-processing steps. Support Vector Machine (SVM) [23], Decision Tree [24] and MLP like algorithms are used to classify extracted features from documents and to identify table regions as well as the table structure.

Zuyev [25] segmented table structure as table grid in rows and column format given the table image as an input. Styles of a table grid are defined based on the visual features and called as a model of a table layout. Considering the presented styles, a table grid structure is detected. When there are no separator lines, rows and columns are detected by analysis of vertical and horizontal profiles, respectively. Intercolumns gap is utilized along with k-means clustering algorithm to identify columns. Various possible separation cases are introduced to identify rows, and heuristics are defined in case of the absence of separators.

Pyreddy and Croft [26] proposed TINTIN to retrieve information from tables by utilizing structural information. The first step is extracting table data and labelling table components as caption, headline or table entries via document preprocessing module. Table extractor makes use of aligned white spaces, characters by considering parameters such as the number of blanks between columns, number of adjacent blanks per line. Then based on many heuristic rules, components of tables are tagged. This process is completed by using syntactic features rather than semantic features. In other words, heuristic rules investigate how components look instead of what they possibly mean. Results proved that tables could be recognized by using visual features along with heuristics.

Kieninger and Dengel [27] presented T-Recs system that has a bottom-up approach based on the clustering word bounding boxes considering their neighbour word bounding boxes. This approach restricts its search to adjacent words in the previous line and the following line. It does not consider separators such as white space that might be very narrow to catch in some cases. Ordered word entries are aggregated as blocks, and segmented

blocks are post-processed to distinguish merged and split columns. River of white spaces, columns with white spaces at the same horizontal position throughout the whole block, and isolated words with neither upper nor lower adjacent word bounding boxes are some of the processes handled in the post-processing step. One advantage of this system is the independence of separators such as white spaces, ruling lines.

TARTAR [28] is aim to construct table from Document Object Model (DOM) tree given the HTML input. It first detects syntactic structure from extracted table data, following noise cleaning, canonicalization and updating DOM trees processes. Then, the Functional Table Model (FTM) rearranges the cell positions while assigning their functional role and semantic labels. In other words, TARTAR extracts structures based on their physical, functional, and semantic features in contrast to [26].

Wang et al. [29] introduced a probability optimization-based algorithm for table structure understanding. In this system, an input image with segmented line and word entities is fed into the algorithm. In the beginning, a column-style labelling algorithm identifies the column style in the document and updates line and word segmentation accordingly. Table candidates are discovered by using background analysis of horizontal blank blocks. A statistical refinement module refines the table detections by doing a vertical projection analysis for the word bounding boxes. The table detection problem is converted to an optimization problem by estimating the probabilities of table entities and text block segmentation outputs. Finally, the table decomposition algorithm constructs columns and extracts cell structures with their attributes (starting/ending row and column). This process is done by using vertical projection on word components similar to the recursive X-Y cut algorithm [30]

Oro and Ruffolo [31] proposed PDF-TREX that heavily relies on PDF metadata and uses a heuristic approach. It has a bottom-up strategy that starts from line tagging and segmenting words in lines using a clustering algorithm based on the whites space threshold. After blocks, rows, columns and table is built in order, content is extracted with cell grid information and outputted in XML format. In another work [32], tables are grouped according to the presence of ruling lines and processed through the bottom-up pipeline of structure recognition rules similar to [31]. Many other tools that rely on PDF metadata can be found in [33].

Kasar et al. [34] presented a work that detects tables by on document images using machine learning algorithms along with the heuristics. The horizontal and vertical lines are extracted from an input image, and a connected component analysis is performed on the extracted image. This analysis consists of 26 defined features that are fed into the machine learning classifier. The trained SVM algorithm classifies regions as table or non-

table by validating each set of intersecting vertical and horizontal separators. Prior to table detection, a simple structure detection is performed based on the grid structure of line separators. However, tables where ruling lines do not present is not considered in the approach.

Ng et al. [35] addressed the table recognition problem under three subproblems. Each subproblem is addressed as a classification problem that one following each other. A decision tree algorithm determines each horizontal line as positive (negative) if the line is inside (outside) of the table. Vertical lines are classified into five defined classes to identify column regions. The last step is to classify horizontal lines inside the table as one of the two classes for row detection. The start/end of row and column is defined easily since transitions from one to another point out these relations. Decision tree and backpropagation algorithms are used to classify entities in a sequential manner similar to followed approaches in recent works. However, the approach solely depends on the presence of line separators.

Rashid et al. [36] proposed a learning-based table recognition framework from heterogeneous documents. The system includes OCR, feature extraction, MLP training and postprocessing steps. After words are obtained using OCR, features are defined according to the distance of a word from its neighbour word components. An autoMLP model is trained to classify each word bounding box as a table or non-table element in the following step. For the recognition of the table, rows are identified using already obtained word bounding box coordinates based on their vertical arrangement. However, the proposed approach is limited to detecting tables and rows.

2.2 Deep Learning Approaches

Deep learning techniques have been adopted widely with their growing popularity and effectiveness in automated document processing [37]. The victory of AlexNet [38]in the ImageNet [39] competition was a watershed moment that paved the way for the creation of more complicated and advanced algorithms in computer vision field. With the advent of CNNs and high processing power from GPUs, many frameworks were proposed for object detection and instance segmentation. Various competitions held on document analysis such as Page Object Detection(POD) [40] propelled deep learning applications to be adopted in a range of tasks from formula detection to tabular region detection. In these analysis, tables and table structures are treated as objects, and the problem is defined as an object detection problem. This thesis focuses on works that address the table structure detection problem alone and adopt CNN-based solutions. Previously the studies that have been done

have focused more on the table detection problem as compared to the structure detection. The reason is that deep learning models are data-driven, and the field suffers from the lack of annotated data with structure information. However, large-scale datasets have been released with ground truth information very recently. We consider deep learning separately when examining machine learning under traditional methods, although deep learning is a sub-branch of machine learning.

DeepDeSRT is one of the first studies to exclude heuristic approaches entirely and depends solely on deep learning. Schreiber et al. [41] propose a deep learning based model for Faster-RCNN-based table detection and Fully Convolutional Network (FCN)-based structure recognition to segment the rows and columns in detected table regions. Additionally, to solve the poor row detection performance due to the proximity of row objects' locations, images are vertically stretched, and bounding boxes that cover less than 0.5% of the input images are removed as pre-processing and post-processing.

Siddiqui et al. [42] propose to use a combination of deformable convolutional networks with Faster R-CNN and FPN. With ResNet-101 as a deformable base model, deformable convolutions solve the issue of fixed receptive field by using extra offsets. These offsets enable the network to change its receptive field depending on the position and object of each input. The authors make datasets that have been annotated with row and column information publicly available.

Paliwal et al. [43] propose utilizing table region information and column information in the encoder level since both table and column have common regions by using FCN architecture with VGG-19 as the backbone. At the decoder level, branches are separated for table and column detection. After table regions and columns are determined, rows are extracted by using Tesseract OCR with heuristic rules.

The authors in [12] empower Mask-RCNN to detect rows and columns. In an extended version of that work [13], a holistic system is presented with the following components: table detection, structure detection and an end-to-end table and structure detection model with an additional judging mechanism for validation of table detections. Prasad et al. [14] present CascadeTabNet that detects tables with their types as bordered and borderless by utilizing Cascade-Mask-RCNN with High Resolution Network (HRNet). The Deep Learning model is used to accomplish cell structure recognition for borderless tables, whereas the line detection algorithm is used for bordered tables with post-processing in both branches. Jiang et al. [15] presented a benchmark on ICDAR 2013 dataset for cell structure detection by using state-of-the-art object detectors in combination with different backbones. Hybrid Task Cascade (HTC) with CBNet double ResNeXt101 outperformed the compared models.

Qiao et al. [44] developed a table structure recognition model by utilizing local texture

and global layout information from the table structure. At the local level, text regions in the tables are aimed to be detected. The challenges here are that the vertical and horizontal spaces between cell regions, empty cells and spanning cells multiple rows/columns. For this reason, the authors obtained the aligned bounding boxes by using global grid boundaries information, which comes from the table layout, grid lines. Aligned bounding boxes are obtained according to the maximum height/width of the non-empty cells in the same row/column. In both Local and Global Pyramid Mask Alignment branches, binary segmentation tasks, which are the same as the original Mask-RCNN model, and pyramid mask regressions with soft-labelled pixels in horizontal and vertical directions are learned simultaneously. Aligned bounding boxes are refined by using a mask re-scoring strategy. Finally, the table structure is recovered by the cell matching, empty cell searching and empty cell merging processes.

Tensmeyer and Martinez improved the SPLERGE [45] method that is composed of two deep learning models. The sub-networks, Row Projection Network (RPN) and Column Projection Network (CPN) of split model outputs r and c the probability of each row (column) of pixels are part of a separator region. Both sub-networks are built on top of a Shared Fully Convolutional Networks (SFCN) that extracts local feature maps. The second model, the Merge model, is designed to recover spanning cells. The input image and outputs of the Split model are fed into the Merge model and returns the probability matrices for up-down and left-right cell merging.

Wei et al. [46] proposed to detect table structure in cell format by utilizing global row and column information. The authors developed a table projection module (TPM) inspired by row pooling and column pooling operations in [45]. TPM is designed to improve table cell object detection by capturing global information instead of detecting row and column separators. TPM module is deployed in the downsample shortcut of ResNet backbone. Zou and Ma [47] localized the row separators, column separators and cell contents by using semantic segmentation. U-Net and DeepLab v2 architectures are used for the semantic segmentation task. Then, a post-processing procedure is applied utilizing bounding boxes of localized identifiers and relative position of row/column separators.

Raja et al. [48] recognize table structure using top-down and bottom-up cues. In the top-down process, cell objects are detected, whereas, in the bottom-up process, tables are rebuilt from cell detections with their row and column relations. The authors modified Mask-RCNN for cell detection and augmented RPN with dilated convolutions to better represent long-range row/column visual features. The top-down pathway structure of FPN is extended to preserve high-level semantic information in low-level feature maps. Additionally, extra loss functions are defined for the definition of row/column relations and alignment of cells. The bottom-up structure recognition process used three different

components based on graph neural networks to catch interactions between the cells and define row/column associations. TabStruct-Net is trained jointly with the cell detection and structure recognition networks to predict cell bounding boxes, row and column adjacency matrices, and cell bounding boxes together.

Zheng et al. [49] proposed a GTE framework that utilizes cell structure detections as a constraint in the table detection task. The presence of cell bounding boxes inside or outside of a predicted table boundary is used to penalize irrational table detections. In the structure recognition task, tables are classified according to their style based on the presence of ruling lines by an attribute network. Then, classified tables are fed into different models which specialize in related styles. Cell bounding box detections are clustered to define row and column locations by the K-means clustering algorithm. Finally, post-processing is performed to split or merge cell bounding boxes.

Hashmi et al. [50] proposed to detect rows and columns separately by optimized anchor generation in the RPN. Anchors are a set of rectangular bounding boxes with predetermined scales and aspect ratios. The anchor generation process for the general object detection task was optimized for the row and column detection model by the authors. A post-processing approach based on black pixels limitations is presented to address the poor row detection problem.

The authors in [51] proposed an attention-based Encoder-Dual-Decoder (EDD) architecture. The encoder is a convolutional neural network (CNN) that extracts the feature map of input table images. The first decoder identifies table structure in row/column format, while the second decoder captures the content of a table cell. Additionally, the authors released a large-scale publicly available table recognition dataset PubTabNet, along with a new evaluation metric Tree-Edit-Distance-based Similarity (TEDS).

Smock et al. [52] presented a large-scale dataset for table extraction, which includes table detection and table structure recognition subproblems along with a new evaluation metric. Grid table similarity (GriTS) provides the flexibility of evaluating table structure recognition from multiple perspectives and across different forms of output. For instance, cell detections can be evaluated by considering 1- layout of the cells, 2- both layout and text content, and 3- both layout and cell locations. As an object detection model, Detection Transformer (DETR) is applied in both table detection and table structure recognition tasks.

Chapter 3

Deep Learning Models for Tabular Structure Recognition

Tables are the most common way to display and communicate structured data. Information stored in tables is of vital importance for a range of organizations, from industry to government. Hence extraction of tabular data in an automated manner is a popular subject in document processing. Obtaining information from tables is carried out in two stages: 1) table detection, 2) table structure recognition. The problem of table detection has been extensively researched and well-studied. However, table structure recognition is a more complex problem to tackle owing to the variety of table kinds and the intricacy of table structure. Deep learning-based solutions have shown promise for table structure recognition, yet further improvement is still needed. To simplify the complex nature of the problem, we adopted deep learning approaches along with analyses that redefine the problem. This section delves into deep learning models that are used to detect table structure in depth.

3.1 Region-Based Convolutional Neural Networks

Girshick et al. [53] developed a method that revolutionized object detection, namely Regions with CNN (R-CNN). According to the previous top result on VOC 2012 [54], the proposed method exceeded the mean average precision (mAP) by roughly 30 percent. Prior to R-CNN, the most successful solutions were ensemble systems that incorporate several low-level image features with high-level context. Girshick et al. demonstrated that CNNs

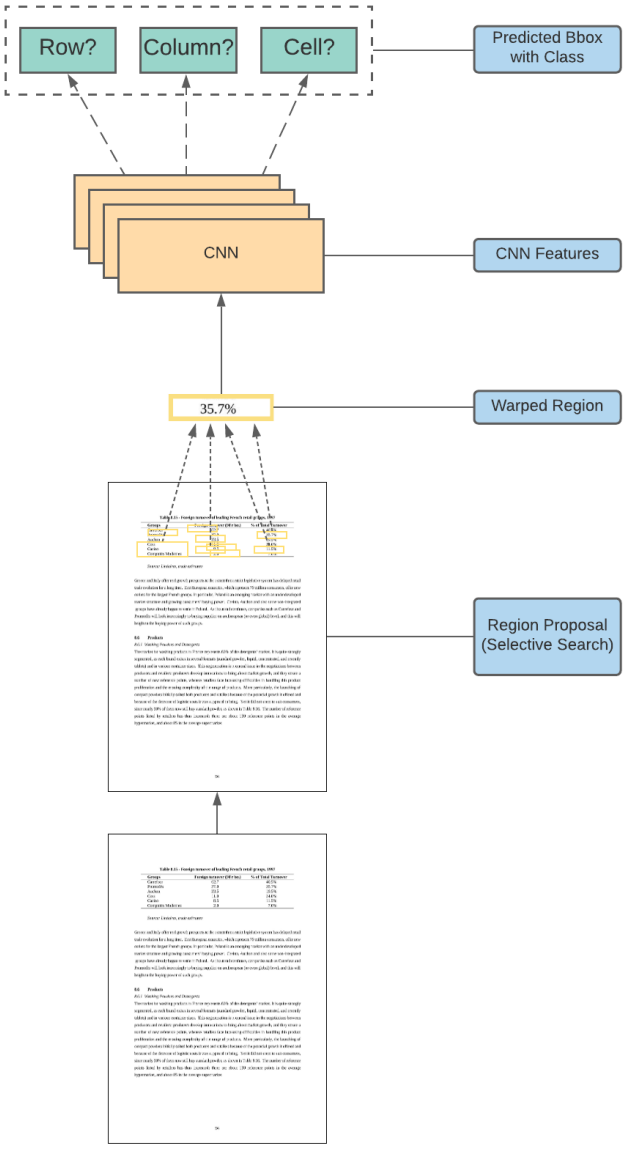


Figure 3.1: R-CNN architecture

undoubtedly can outperform SIFT and HOG-like features in object detection performance. This method has three key modules. Firstly, a high number of region proposals are generated through a selective search approach which can be replaced by any other method.

Those are the regions where an object might present on the image, and they are class independent. The region proposal method realizes object detection with CNN more efficiently than the sliding window approach that exhaustively applies CNNs to the particular image area until the entire image has been checked and classified. The second module is feature extraction, where features are extracted from each region proposal in a fixed length of a vector by a CNN following the implementation of AlexNet [38]. All region images are converted to the fixed input size through a warping process of pixels. At this point, another contribution of [53] plays a crucial role in detection performance. The authors showed the effectiveness of transfer learning by using pre-trained CNNs on a large scale ImageNet dataset [55] to fine-tune on a small dataset (PASCAL). This makes it possible to develop domain-specific object detection models even when the labelled data is scarce. The third module is to classify extracted feature images. Here, linear SVMs, which each specializes in a specific class, are used for classification. Although R-CNN [53] is not utilized in this thesis, a flowchart for structure detection using R-CNN is provided in Figure 3.1 to help readers comprehend the used object detection and instance segmentation algorithms.

[53] spawned the class of two-stage object detectors. As explained above, it is the additional component of the region proposal network prior to feature extraction. Two-staged region-based CNN algorithms are widely employed in object detection [56, 57, 58] and have been applied to the problem of detecting table structure. Faster R-CNN has been shown to perform well, and various enhanced versions are proposed for use in combination with advanced backbone architectures.

3.1.1 Fast(er) R-CNN

The Region-based Convolutional Network (R-CNN) gave a new impulse to object detection accomplishing excellent accuracy. However, it still suffers from slowness in training and inference steps and high computational cost in memory and time. Girshick [17] addressed these issues and proposed the Fast R-CNN method. It takes image and region proposals as inputs. The input image is processed to extract a feature map, and then a fixed-size feature vector is created by the region of interest (RoI) pooling layer for each region proposal. The RoI pooling layer uses max pooling to create a feature vector from the features inside the region of interest. The difference here is that the image is processed in one pass instead of feeding each region proposal to CNN. It is a very crucial aspect in the speed of training, especially when considering the 2000 region proposal per image. The resultant feature vector is successively sent to two fully connected (fc) layers that finally branches into two layers. One of the branches outputs the softmax class probabilities for each RoI, $p = (p_0, ..., p_K)$ for $K + 1$ categories the plus value being the background. The

other branch, the bounding-box regressor, outputs the offsets $t^k = (t^k_x, t^k_y, t^k_w, t^k_h)$ for $k \in K$ bounding-box information.

A part of the efficiency of Fast R-CNN stems from the computation and memory sharing during the training. The features of RoIs from the same image are shared by the hierarchical sampling of images in mini-batches and RoIs in the forward and backpropagation steps. The proposed scheme provides 64 times faster training time. Also, the method proposes joint training of softmax classifier and bbox regressor, unlike separate training stages of a softmax classifier, SVMs, and regressors in [53]. To achieve a combined training pipeline for classification and bounding-box regression, a multi-task loss function is defined where u and v is the ground-truth class information and ground-truth bbox regression location.

$$L(p, u, t^u, v) = L_{cls}(p, u) + \lambda[u \geq 1]L_{reg}(t^u, v) \tag{3.1}$$

The term $L_{cls}(p, u) = -\log p_u$ is the logarithmic loss of the actual class u. The L_{reg} is the loss of the real bbox regression target over the predicted location where both in a tuple format. The loss used for bbox regression:

$$L_{reg}(t^u, v) = \sum_{i \in x,y,w,h} smooth_{L_1}(t^u_i - v_i), \tag{3.2}$$

$$smooth_{L_1}(x) = \begin{cases} 0.5x^2 & if|x|< 1 \\ |x|-0.5 & otherwise, \end{cases} \tag{3.3}$$

The author used L_1 loss because it is less sensitive to variations and outliers compared to L_2 loss used in R-CNN.

The region proposal procedure is time-consuming and computationally expensive in state-of-the-art object detection networks. The Selective Search method adapted by [17] spends two seconds per image to generate region proposals by running on a CPU. Fast R-CNN improved the speed and accuracy remarkably; however, there is still much room for growth. Ren et al. [18] propose a new algorithm, Region Proposal Network (RPN), for region proposal that runs on a GPU. RPN is a deep fully convolutional network that shares convolutional layers with the object detection network, Fast R-CNN. These two networks are unified through various training schemes by fine-tuning for region proposal and object detection. The new approach, RPNs with Fast R-CNN that constitutes Faster R-CNN, reduces the running time to 10 milliseconds per image for the region proposal.

The RPN takes feature maps as input and outputs rectangular region proposals with an objectness score that measures a region belonging to a class of objects or background.

RPN is a fully convolutional network that has common layers with the feature extractor of the object detection network, allowing the two networks to share information. A small network slides over the feature map from the final shared convolutional layer to generate proposals. The input size is $n \times n$ to the small network. The resulting output is a 512 dimension feature vector. The feature vector is sent to fully connected layers for bbox regression and bbox classification. Also, a new so-called "anchor" method is introduced to generate region proposals at multiple scales and aspect ratios, unlike randomly sized RoIs. Multiple region proposals are predicted for each sliding-window location where the maximum number of proposals is defined as k. Hence box regression layer yields $4k$ outputs with coordinates, and the box classification layer produces $2k$ objectness scores. Anchors are parameterized proposals in relation to k reference boxes. They have three different scales and aspect ratios, resulting in $k = 9$ anchors at each position. There are $W \times H \times k$ anchors in total for a feature map with a size of $W \times H$.

Three different ways are identified for training RPN and detection networks with features shared rather than separate learning. The adopted strategy is Four-Step Alternating Training. First, the RPN is trained with ImageNet-pre-trained weights for region proposal task. In the second step, Fast R-CNN as the detection network is trained using the first step proposals and initialized in the same way. So far, there has been no information sharing between the two networks. In the third step, RPN training is initialized with the detector network, and only layers unique to RPN are fine-tuned while shared layers are frozen. Finally, the detection network's particular layers are fine-tuned, while the shared layers are fixed. The unified network is realized with information sharing through common convolutional layers. The multi-task loss function in Fast R-CNN is adopted with the addition of binary class labels that is object or not and defined as

$$L(p_i, t_i) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*). \tag{3.4}$$

p_i is the predicted probability of ith anchor, and p_i^* is the target value which is either 1 or 0 depending on whether the anchor is positive. If the anchors IoU overlapping is greater than 0.7 it is labelled as positive. A negative label is assigned when the IoU ratio is smaller than 0.3. Anchors without positive or negative labels are not used in training. The classification loss L_{cls} is the logarithmic loss of two classes object or not object as in the Fast R-CNN. The regression loss is also adopted from Fast R-CNN and defined as $L_{reg}(t_i, t_i^*) = smooth_{L_1}(t_i - t_i^*)$. The p_i^* coefficient of the L_{reg} is to penalize negative anchors contribution to regression loss. The classification and regression loss are normalized by N_{cls} and N_{reg} that are mini-batch size and the number of anchor locations, respectively.

3.1.2 Mask R-CNN

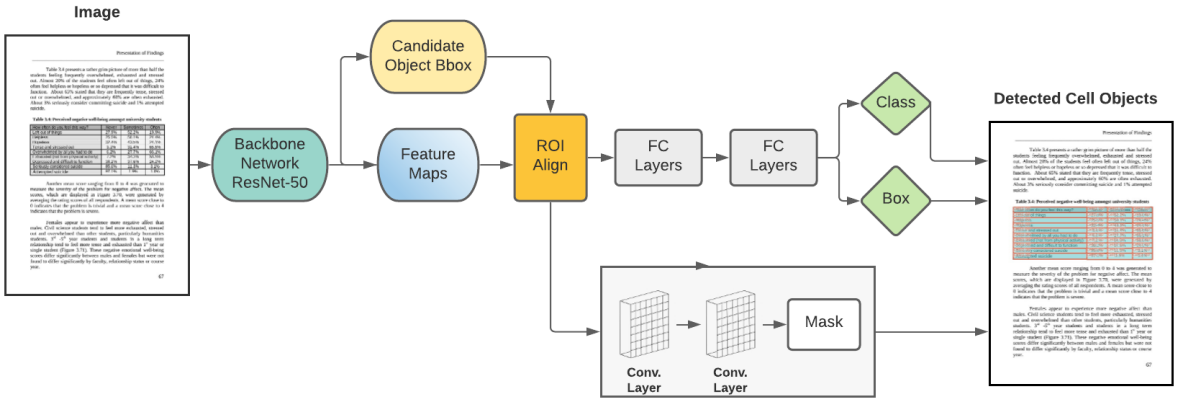


Figure 3.2: Mask R-CNN architecture for table cell structure detection

The R-CNN, Fast R-CNN and Faster R-CNN structures are explained in detail in the previous sections. Understanding how these detectors work makes it significantly easier to understand the methods used in this thesis since they constitute the foundation for two-staged object detectors. Mask R-CNN is an instance segmentation framework that is built with the addition of a mask branch to Faster R-CNN by He et al. [19]. The goal here is to segment each instance of objects in a pixel-wise manner in addition to the bbox classification and localization. The added mask branch is a small Fully Convolutional Network (FCN) that predicts segmentation masks for each RoI in pixel level. The first stage, same as in [18] generates proposal bounding boxes where objects possibly lie via Region Proposal Network (RPN). In the second stage, features are extracted from each region proposal using RoIPool to output class prediction and box location. Classification and regression of bounding-box stage were adopted from Fast/Faster R-CNN. However, Mask R-CNN uses RoIAlign instead of RoIPool [17] for feature extraction from RPN outputs. RoIPool causes misalignment between the RoI and extracted features due to the quantization process. RoIAlign solves this problem with a simple change in quantization and aligns the extracted features with the input.

The multi-task loss function used in training defined as $L = L_{cls} + L_{reg} + L_{mask}$. The classification loss L_{cls} and regression loss L_{reg} are adopted from [17] and same as defined in Equations 3.1 and 3.2, respectively. The mask branch outputs K binary masks of size $m \times$ m for each RoI out of K classes. Hence the resulting output is Km^2 dimensional per

RoI. The mask loss L_{mask} is defined as the average binary cross-entropy loss with sigmoid activation for each pixel. The mask prediction is class agnostic, and the prediction of box labels is made in the dedicated classification branch. The separate mask and class prediction prevent mask branch competing among classes that is the key factor in effective instance segmentation.

The RoIPool extracts features from RoIs and collapses them into vectors by quantization. The quantization of RoI features results in a mismatch of the RoI and extracted features. This process is initially designed for box classification and regression and is not suitable for pixel-wise prediction. The prediction of mask objects for each RoI requires a high-quality feature map that preserves the spatial layout, unlike the prediction of box location and class. To address this issue, He et al. introduced the RoIAlign layer that replaces RoIPool. The new method uses bilinear interpolation instead of quantization and provides better alignment between extracted features and input. The effect on mask accuracy returns as a 10% to 50% improvement is noteworthy. Fig. 3.2 illustrates the flow of the proposed Mask R-CNN method in this thesis.

3.1.3 Cascade Mask R-CNN

As we mentioned in Section 3.1.1, a value of IoU threshold is used to assign positive or negative labels in object detection. Using a single IoU threshold value, e.g. 0.5, in training causes noisy proposals that lead to false-positive detections. Distinguishing the backgrounds and objects with such a value of IoU threshold is a loose constraint because proposals that people perceive as false positives exceed the defined threshold value, which is commonly 0.5. Increasing the IoU threshold, on the other hand, have a significant negative impact on network performance. Two reasons are put forward for the paradox of high-quality object detection. First, increasing the Iou threshold causes the exponential decrease in the distribution of positive samples, and the overfitting problem appears. Second, the IoU value mismatch at training and inference time. The detector is optimized for a particular IoU value over the proposals during training and performs poorly in other IoU thresholds. To address these issues, Cai et al. [59] propose Cascade R-CNN that is the extended version of Faster R-CNN with the sequential design of detection head where class and box predictions are made. The output of one stage is fed as input to the next step to adjust box predictions instead of eliminating negative samples. This cascaded structure refines predictions with a resampling mechanism at each stage helps to solve the overfitting problem by keeping a set of close false positives for detectors adapted higher IoUs in the latter stages. Also, applying the same sequential procedure at inference time

better matches the gradually improved proposals and detectors that have increasing IoU thresholds.

The loss function for bounding box regression is identical to [18] and defined as $L_{reg}(f(x_i, b_i), g_i)$. The candidate bounding box $b = (b_x, b_y, b_w, b_h)$ that has the coordinate information is regressed over the ground truth bounding box g with a regressor $f(x, b)$. The index i indicates the proposed sample. However, multi-stage object detectors requires a progressive design of regressor that is optimized for multiple IoU values and different proposal distirbutions due to explained reasons above. Hence, authors designed a sequential regressors

$$f(x, b) = f_T \circ f_{T-1} \circ \cdots \circ f_1(x, b) \tag{3.5}$$

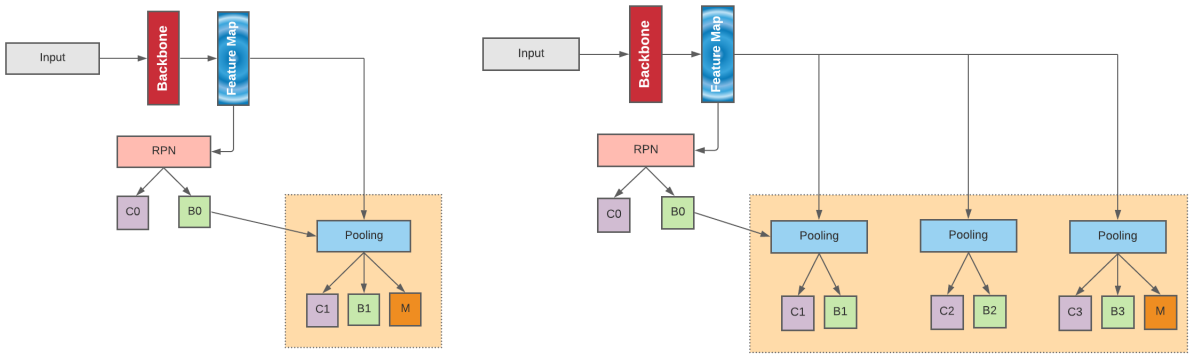
where the output of one stage is fed as input to the other. T is the number of total stages and each f_t is optimized for the arriving sample distribution b^t rather than the original distribution b^1. Also, to increase the number of positive samples, they are resampled for higher IoU values at each stage. The overall loss function for Cascade R-CNN is

$$L(x^t, g) = L_{cls}(h_t(x^t), y^t) + \lambda[y^t \geq 1]L_{reg}(f_t(x^t, b^t), g), \tag{3.6}$$

$$L_{cls}(h_t(x^t), y^t) = \sum_{u \in U} L_{cls}(h_u^t(x_u^t), y_u^t) \tag{3.7}$$

where U is a set of increasing IoU threshold values and h_t and f_t are optimized classifier and regressor for IoU value of u_t at stage t.

Cai et al. extended Cascade R-CNN for instance segmentation and proposed Cascade Mask R-CNN [20]. It is built with the addition of a mask branch in parallel to the detection branch, similar to Mask R-CNN. However, Cascade R-CNN has multiple detection branches, and the issue is determining where to place mask branches and how many of them. So, authors in [20] investigated different strategies where the segmentation branch is added only to the first stage, last stage or all stages. Trained mask head in the first stage causes a mismatch between the mask predictions since inference is made after the third stage. The number of positive samples varies at each stage, and the final stage has more instances. For this reason, adding the mask head to the last stage requires pixels-wise prediction of highly overlapping examples that is not as helpful as in object detection. The alternative approach is to add a mask branch to all stages and obtain the final mask prediction from a group of mask branches. This approach adds extra computation and memory but does not provide a noticeable improvement over the strategy that has a mask branch only at the final stage. Hence, we followed the approach that has a mask head only at the final stage. The design of our Cascade R-CNN implementation is given in Figure 3.3b along with Mask R-CNN 3.3a.



(a) Mask R-CNN

(b) Cascade Mask R-CNN

Figure 3.3: The architecture of Mask R-CNN (a) and Cascade Mask R-CNN (b). C, B and M denote class label, bounding box and mask predictions of the corresponding stage, respectively. Proposal head outputs are denoted with zero indexes.

3.1.4 Hybrid Task Cascade

Cascaded design of multiple detectors and addition of mask branch for instance segmentation improved the performance of Mask R-CNN. The detection and mask heads, on the other hand, are optimized independently and do not benefit from one another. The only advantage is that mask branches segment the refined bounding boxes in the latter stages. Since there is no information flow between them, there is no direct contribution. Hybrid Task Cascade (HTC) [21] was proposed to improve Cascade Mask R-CNN by exploiting the relationship between the detection and segmentation heads. HTC interleaves the box and mask branches instead of performing bounding box predictions and mask predictions in parallel branches. Thus each mask head utilizes the refined bounding box predictions from the next layer. Also, HTC introduces an information flow between the mask predictions at different stages. Mask features at each stage are fed to the next mask branch directly. Another contribution is that a fully convolutional semantic segmentation branch is added in HTC architecture. The segmentation branch provides additional features obtained from Feature Pyramid [60] to the box and mask heads of each stage. This information-sharing design of architecture among the detection, mask and semantic segmentation heads is the key to improvement.

The intertwining design of the box, mask and semantic segmentation branches is formulated as

$$
\begin{aligned}
x_t^{bbox} &= P(x, r_{t-1}), \quad & r_t &= B_t(x_t^{bbox}), \\
x_t^{mask} &= P(x, r_t), & m_t &= M_t(F(x_t^{mask}, m_{t-1}^{-})).
\end{aligned}
\tag{3.8}
$$

x_t^{bbox} and x_t^{mask} represents the bounding box and mask features obtained from CNN features x and RoIs. At any stage t B_t and M_t are the detection and segmentation heads, while r_t and m_t are the corresponding bounding box and mask predictions. $P(\cdot)$ indicates a pooling operator such as RoIPool or RoIAlign. Here, it can be seen that instead of both bbox and mask branches are fed by the previous stage, mask head take the refined bounding boxes as input. The mask feature of the current stage x_t^{mask} is merged with the intermediate feature m_{t-1}^{-} from the previous stage M_{t-1} using the F function to accomplish information flow between mask branches. The flow of HTC is represented in Figure 3.4 with the same convention in Figure 3.3.

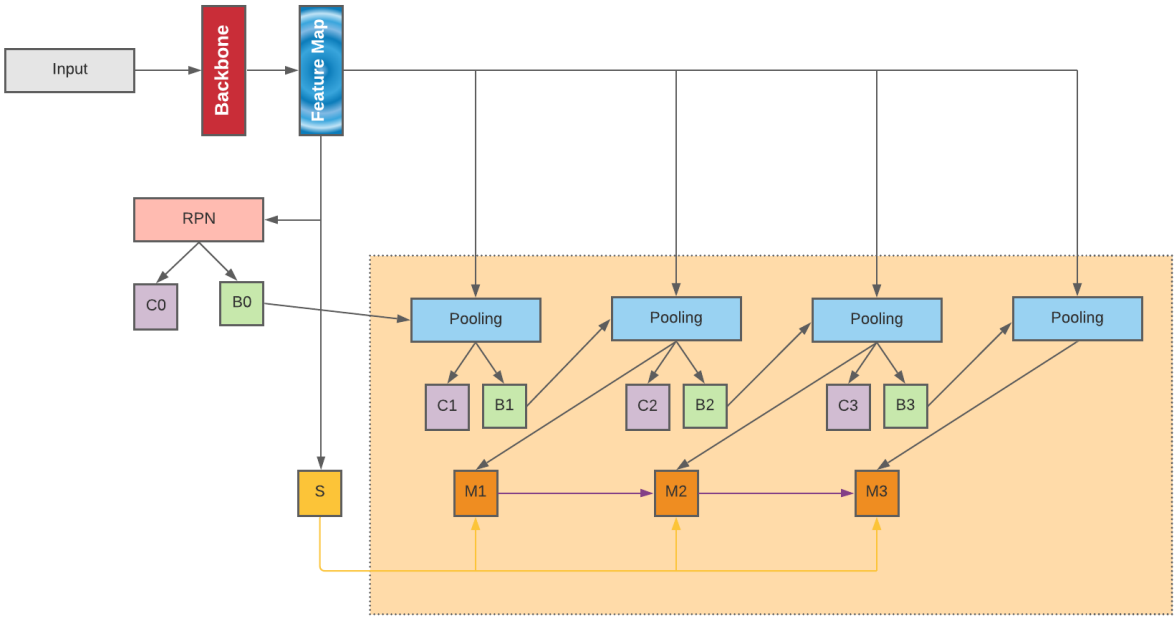


Figure 3.4: Hybrid Task Cascade architecture. S component represents the Semantic Segmentation Branch.

3.2 Backbone Networks

Region-based CNNs revolutionized object detection and instance segmentation with their performance and efficiency. The critical component to achieve this is the RPN that provides region proposals. Both region proposal and detector networks utilize the feature map of an input image. Backbone networks operate as feature extractors and generate these feature maps. AlexNet[38], VGG [61], GoogLeNet [62], ResNet [63], are the first effective deep learning-based CNN approaches that made an impact on ImageNet dataset for the image classification task. With the transfer learning paradigm, they are used as backbone networks to extract features in object detection and instance segmentation. The need for deeper networks for higher accuracy led to the design of more sophisticated backbone networks. We utilize these advanced backbone networks, e.g. ResNet [63], ResNeXt [64], HRNet [65] and CBNet [66] and explain them in the following sections.

3.2.1 ResNet

The performance of [38] on image classification ignited the design of deeper networks. However, simply stacking more layers does not provide better learning. One reason for this is the well-known problem of the vanishing gradient. This problem has been resolved mainly by the normalization process in the initialization and intermediate layers. Another issue is degradation, which arises as deep networks begin to converge. The accuracy saturates and drops dramatically as the depth increases. He et al. [63] proposed a deep residual learning framework to address these issues. Unlike plain networks, the stacked layers optimize the residual form of original mapping. Given a feature map $h(x)$, the residual equation is $F(x) := H(x) - x$ and the resulting feature map is $F(x) + x$. This is accomplished by shortcut connections, which skip a few levels and forward the output from the previous layer to the one ahead. If the dimensions are not the same, x is multiplied by linear projection W_s to match the sizes with F otherwise added element-wise. The authors extended 34-layer ResNet to 50, 101 and 152 layers with additional blocks and still have less complexity than VGG-16/19. On the other hand, the accuracy is remarkably improved on the COCO [67] dataset with ResNet-101 compared to VGG-16. We implemented both ResNet-50/101 in this study.

3.2.2 HRNet

In typical frameworks such as AlexNet, VGG and ResNet, images are downsampled from high resolutions to low resolutions to obtain feature maps. Then an upsampling

process is applied to construct high-resolution feature maps from low-resolution representations. These processes are done by convolutional operations connected in series. The convolution and deconvolution operations cause the loss of spatial information and provide semantically poor feature maps. To address these issues, Wang et al. [65] proposed two versions of high-resolution networks (HRNetV1 and HRNetV2). HRNetV2 is designed in a multi-stage manner where high-resolution representations are kept, and one lower resolution representation is added at each step in parallel. Those added low-resolution feature maps boost the high-resolution representations rather than directly being recovered. Besides, information is exchanged between different resolutions by a fusion module that provides semantically rich feature maps with more precise spatial information.

3.2.3 CBNet

Composite Backbone Network (CBNet) [66] introduced the idea of composite design of existing backbone networks such as ResNet [63] and ResNeXt [64]. There are two types of backbones, namely the Lead Backbone and the Assistant Backbone in CBNet. The assistant backbone refines the leading backbone outputs at each stage, and refined features are fed into the following stage of the leading backbone. Backbones are called as Dual-Backbone or Triple-Backbone depending on the number of identical backbones assembled.

Chapter 4

Table Structure Ground-Truthing

As seen in Chapter 2, it is proved that the state-of-the-art deep learning-based solutions are great candidates to detect table structure. However, deep learning models are data-driven solutions and require annotated data with ground-truth information. The challenges here are two-fold. First, hand labelling images with table structure information is a tedious, time-consuming and expensive process. Second, accurate ground-truthing of table structure is problematic because the truth may exist in more than one acceptable form. Also, the format in which the structure is defined is critical. In this chapter, table structure detection is analyzed under different structure formats such as row, column and cell. Following the discovery of the optimal format, an efficient table structure labelling technique for annotation bootstrapping is proposed.

4.1 Table Structure Definition

Table structure can be defined in different formats such as row/column and cell. Each definition requires a unique strategy to detect table structure and recognize the relations among the defined table elements. However, ground-truthing is a complicated process due to the variety of table types and the complexity of table structure. During our analyses, we have discovered some reasons that explain the difficulties behind accurate ground-truthing. In various instances, human operators may have differing viewpoints on the structure's correct truth, and there might be several acceptable ways to define them. Inconsistencies are unavoidable when multiple human operators realize the annotation process while even the same person may complete inconsistent ground-truth annotations in the same dataset. It is very likely that erroneous and inconsistent labelling causes performance degradation in

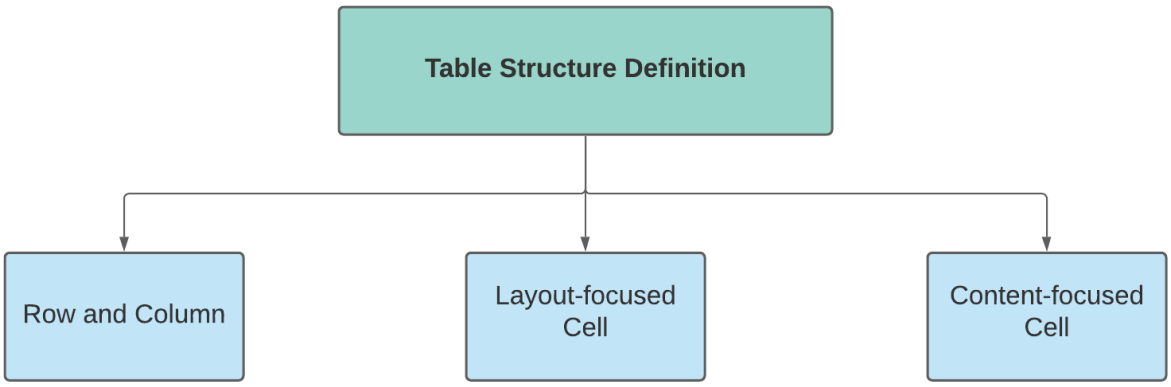


Figure 4.1: Table Structure Definition

deep learning models. Another critical point is the presence of ruling lines. When there is no existing ruling line, it is very hard to establish the boundaries of table elements. Considering the aforementioned reasons and the ambiguities ground-truthing table structure is complicated and has a critical effect on the performance of data-driven models.

The following subsections analyze the performance of the deep learning model for row/column and cell structure formats. The taxanomy of table structure definition is given in Figure 4.1. Beyond the performance, the advantages, disadvantages and challenges are investigated to find the optimal definition of table structure.

TABLE 3. Number of deaths and age-specific death rates* for coronary heart disease among adults aged ≥45 years, by age group, sex, and Hispanic ethnicity — National Vital Statistics System, United States, 2006

Age group (yrs)	Women						Men					
	Hispanic			Non-Hispanic			Hispanic			Non-Hispanic		
	No.	Rate	(95% CI)	No.	Rate	(95% CI)	No.	Rate	(95% CI)	No.	Rate	(95% CI)
45–54	345	15.5	(13.8–17.1)	5,663	28.7	(27.9–29.4)	1,205	52.7	(49.7–55.7)	17,707	93.2	(91.8–94.6)
55–64	806	60.9	(56.7–65.1)	12,273	81.6	(80.2–83.0)	1,906	156.5	(149.5–163.6)	31,564	225.4	(222.9–227.8)
65–74	1,512	199.2	(189.2–209.2)	22,270	234.7	(231.6–237.8)	2,430	394.1	(378.5–409.8)	40,266	500.0	(495.1–504.9)
75–84	3,012	666.6	(642.8–690.4)	54,839	751.6	(745.3–757.9)	3,235	1,022.8	(987.6–1,058.1)	63,916	1,282.9	(1,273.0–1,292.9)
≥85	3,694	2,213.2	(2,141.8–2,284.5)	94,269	2,739.1	(2,721.6–2,756.6)	2,176	2,453.9	(2,350.8–2,557.0)	53,499	3,344.5	(3,316.2–3,372.9)
Total	9,369	190.0	(186.2–193.9)	189,314	344.1	(342.5–345.6)	10,952	242.0	(237.5–246.5)	206,952	434.4	(432.5–436.2)

Abbreviation: CI = confidence interval.
*Per 100,000 U.S. standard population.

TABLE 4. Number of deaths and age-specific death rates* for stroke among adults aged ≥45 years, by age group, sex, and race — National Vital Statistics System, United States, 2006

Age group (yrs)	Race											
	American Indian/Alaska Native			Asian/ Pacific Islander			Black			White		
	No.	Rate	(95% CI)	No.	Rate	(95% CI)	No.	Rate	(95% CI)	No.	Rate	(95% CI)
Women												
45–54	19	—	—	109	10.4	(8.4–12.3)	875	31.3	(29.3–33.4)	1,856	10.4	(9.9–10.8)
55–64	22	16.2	(10.2–24.5)	202	28.8	(24.8–32.7)	1,090	61.1	(57.5–64.7)	3,307	24.1	(23.2–24.9)
65–74	55	78.8	(59.4–102.5)	322	80.8	(72.0–89.6)	1,565	148.9	(141.5–156.3)	6,918	79.3	(77.4–81.1)
75–84	99	267.6	(217.5–325.8)	669	284.2	(262.7–305.7)	2,701	415.6	(400.0–431.3)	21,943	321.5	(317.2–325.7)
≥85	106	648.1	(524.7–771.5)	621	777.0	(715.9–838.2)	2,901	1,060.5	(1,021.9–1,099.1)	35,698	1,102.2	(1,090.8–1,113.7)
Total	301	63.4	(56.3–70.6)	1,923	77.9	(74.5–81.4)	9,132	139.4	(136.5–142.3)	69,722	138.2	(137.1–139.2)
Men												
45–54	33	16.3	(11.2–22.9)	126	13.4	(11.1–15.8)	1,044	43.5	(40.9–46.2)	2,279	12.8	(12.3–13.4)
55–64	44	35.0	(25.4–47.0)	220	36.3	(31.5–41.1)	1,523	105.9	(100.6–111.3)	4,110	31.5	(30.5–32.4)
65–74	50	82.9	(61.5–109.3)	357	108.9	(97.8–120.2)	1,644	218.7	(208.1–229.2)	7,312	97.1	(94.9–99.3)
75–84	48	174.3	(128.5–231.1)	477	294.9	(268.5–321.4)	1,741	471.1	(449.0–493.2)	16,041	338.5	(333.2–343.7)
≥85	27	344.5	(227.0–501.2)	417	865.9	(782.8–949.0)	987	882.0	(827.0–937.0)	14,311	941.3	(925.9–956.7)
Total	202	47.6	(41.1–54.2)	1,597	76.7	(73.0–80.5)	6,939	136.9	(133.7–140.1)	44,053	98.8	(97.9–99.7)

Abbreviation: CI = confidence interval.
* Per 100,000 U.S. standard population.
† Number of deaths too small to calculate a reliable rate.

TABLE 5. Number of deaths and age-specific death rates* for stroke among adults aged ≥45 years, by age group, sex, and Hispanic ethnicity — National Vital Statistics System, United States, 2006

Age group (yrs)	Women						Men					
	Hispanic			Non-Hispanic			Hispanic			Non-Hispanic		
	No.	Rate	(95% CI)	No.	Rate	(95% CI)	No.	Rate	(95% CI)	No.	Rate	(95% CI)
45–54	263	11.8	(10.4–13.2)	2,590	13.1	(12.6–13.6)	389	17.0	(15.3–18.7)	3,080	16.2	(15.6–16.8)
55–64	368	27.8	(25.0–30.7)	4,243	28.2	(27.4–29.1)	501	41.1	(37.5–44.8)	5,380	38.4	(37.4–39.4)
65–74	584	76.9	(70.7–83.2)	8,256	87.0	(85.1–88.9)	617	100.1	(92.2–108.0)	8,723	108.3	(106.0–110.6)
75–84	1,087	240.6	(226.3–254.9)	24,285	332.8	(328.6–337.0)	926	292.8	(273.9–311.6)	17,350	348.2	(343.1–353.4)
≥85	1,240	742.9	(701.6–784.3)	38,056	1,105.8	(1,094.6–1,116.9)	516	581.9	(531.7–632.1)	15,203	950.4	(935.3–965.5)
Total	3,542	71.8	(69.5–74.2)	77,430	140.7	(139.7–141.7)	2,949	65.2	(62.8–67.5)	49,736	104.4	(103.5–105.3)

Abbreviation: CI = confidence interval.
* Per 100,000 U.S. standard population.

as diabetes, which varies substantially across racial/ethnic groups. Finally, state of residence at death from CHD and stroke — diseases that often have long latency periods — might not reflect the location of the decedent's lifetime health, access to health care, and state cardiovascular health promotion activities.

The proposed *Healthy People 2020* objectives for heart disease and stroke were developed to prevent premature death from cardiovascular disease by maintaining low risk for disease, controlling increased risk, detecting and treating heart attacks and strokes, and reducing disability and recurrence (*12*). Research examining health

Figure 4.2: Ground-Truth Annotations for Row and Column Structure

4.1.1 Table Structure Detection in Row and Column Format

Rows and columns are the fundamental elements of tables. Defining tables by using row/column simplifies identifying relations between the key-value pairs. In other words, it helps to understand which value belongs to which key element in the header. However, when tables have a complex structure, it is very challenging to distinguish rows/columns. Heterogeneous tables where spanning rows/columns more than one cell are examples of complex tables. This is a hurdle for both deep learning models and ground-truthing. Figure 4.2 presents an image document with complex tables that are annotated in row and column format as an example. The centre points of the separator spaces are considered as borders when annotating tables that lack ruling lines.

Row objects tend to be long in width but very short in height. Column objects usually have moderate width while long in height. Since row and column objects have different characteristics and the covered area is highly overlapped, they are both trained separately and together for detailed investigation. An example of the separately annotated document image with row and column objects is given in Figure 4.3 and Figure 4.4 respectively.

TABLE 3. Number of deaths and age-specific death rates* for coronary heart disease among adults aged ≥45 years, by age group, sex, and Hispanic ethnicity — National Vital Statistics System, United States

Age group (yrs)	Women						Men					
	Hispanic			Non-Hispanic			Hispanic			Non-Hispanic		
	No.	Rate	(95% CI)	No.	Rate	(95%	No.	Rate	(95% CI)	No.	Rate	(95% CI)
45–54	345	15.5	(13.8–17.1)	5,663	28.7	(27.9–	1,205	52.7	(49.7–55.7)	17,707	93.2	(91.8–94.6)
55–64	806	60.9	(56.7–65.1)	12,273	81.6	(80.2–	1,906	156.5	(149.5–163.6)	31,564	225.4	(222.9–227.8)
65–74	1,512	199.2	(189.2–209.2)	22,270	234.7	(231.6–	2,430	394.1	(378.5–409.8)	40,266	500.0	(495.1–504.9)
75–84	3,012	666.6	(642.8–690.4)	54,839	751.6	(745.3–	3,235	1,022.8	(987.6–1,058.1)	63,916	1,282.9	(1,273.0–1,292.9)
≥85	3,694	2,213.2	(2,141.8–2,284.5)	94,269	2,739.1	(2,721.6–	2,176	2,453.9	(2,350.8–2,557.0)	53,499	3,344.5	(3,316.2–3,372.9)
Total	**9,369**	**190.0**	**(186.2–193.9)**	**189,314**	**344.1**	**(342.5–345.6)**	**10,952**	**242.0**	**(237.5–246.5)**	**206,952**	**434.4**	**(432.5–436.2)**

Abbreviation: CI = confidence interval.
*Per 100,000 U.S. standard population.

TABLE 4. Number of deaths and age-specific death rates* for stroke among adults aged ≥45 years, by age group, sex, and race — National Vital Statistics System, United States, 2006

Age group (yrs)	American Indian/Alaska Native			Asian/ Pacific Islander			Black			White		
	No.	Rate	(95% CI)	No.	Rate	(95%	No.	Rate	(95% CI)	No.	Rate	(95% CI)
Women												
45–54	19	—	—	109	10.4	(8.4–1	875	31.3	(29.3–33.4)	1,856	10.4	(9.9–10.8)
55–64	22	16.2	(10.2–24.5)	202	28.8	(24.8–	1,090	61.1	(57.5–64.7)	3,307	24.1	(23.2–24.9)
65–74	55	78.8	(59.4–102.5)	322	80.8	(72.0–	1,565	148.9	(141.5–156.3)	6,918	79.3	(77.4–81.1)
75–84	99	267.6	(217.5–325.8)	669	284.2	(262.7–	2,701	415.6	(400.0–431.3)	21,943	321.5	(317.2–325.7)
≥85	106	648.1	(524.7–771.5)	621	777.0	(715.9–	2,901	1,060.5	(1,021.9–1,099.1)	35,698	1,102.2	(1,090.8–1,113.7)
Total	301	63.4	(56.3–70.6)	1,923	77.9	(74.5–	9,132	139.4	(136.5–142.3)	69,722	138.2	(137.1–139.2)
Men												
45–54	33	16.3	(11.2–22.9)	126	13.4	(11.1–	1,044	43.5	(40.9–46.2)	2,279	12.8	(12.3–13.4)
55–64	44	35.0	(25.4–47.0)	220	36.3	(31.5–	1,523	105.9	(100.6–111.3)	4,110	31.5	(30.5–32.4)
65–74	50	82.9	(61.5–109.3)	357	108.9	(97.6–	1,644	218.7	(208.1–229.2)	7,312	97.1	(94.9–99.3)
75–84	48	174.3	(128.5–231.1)	477	294.9	(268.5–	1,741	471.1	(449.0–493.2)	16,041	338.5	(333.2–343.7)
≥85	27	344.5	(227.0–501.2)	417	865.9	(782.8–	987	882.0	(827.0–937.0)	14,311	941.3	(925.9–956.7)
Total	202	47.6	(41.1–54.2)	1,597	76.7	(73.0–80.5)	6,939	136.9	(133.7–140.1)	44,053	98.8	(97.9–99.7)

Abbreviation: CI = confidence interval.
* Per 100,000 U.S. standard population.
† Number of deaths too small to calculate a reliable rate.

TABLE 5. Number of deaths and age-specific death rates* for stroke among adults aged ≥45 years, by age group, sex, and Hispanic ethnicity — National Vital Statistics System, United States, 2006

Age group (yrs)	Women						Men					
	Hispanic			Non-Hispanic			Hispanic			Non-Hispanic		
	No.	Rate	(95% CI)	No.	Rate	(95	No.	Rate	(95% CI)	No.	Rate	(95% CI)
45–54	263	11.8	(10.4–13.2)	2,590	13.1	(12.6	389	17.0	(15.3–18.7)	3,080	16.2	(15.6–16.8)
55–64	368	27.8	(25.0–30.7)	4,243	28.2	(27.4	501	41.1	(37.5–44.8)	5,380	38.4	(37.4–39.4)
65–74	584	76.9	(70.7–83.2)	8,256	87.0	(85.1	617	100.1	(92.2–108.0)	8,723	108.3	(106.0–110.6)
75–84	1,087	240.6	(226.3–254.9)	24,285	332.8	(328.6	926	292.8	(273.9–311.6)	17,350	348.2	(343.1–353.4)
≥85	1,240	742.9	(701.6–784.3)	38,056	1,105.8	(1,094.7	9) 516	581.9	(531.7–632.1)	15,203	950.4	(935.3–965.5)
Total	**3,542**	**71.8**	**(69.5–74.2)**	**77,430**	**140.7**	**(139.7–141.7)**	**2,949**	**65.2**	**(62.8–67.5)**	**49,736**	**104.4**	**(103.5–105.3)**

Abbreviation: CI = confidence interval.
* Per 100,000 U.S. standard population.

as diabetes, which varies substantially across racial/ethnic groups. Finally, state of residence at death from CHD and stroke — diseases that often have long latency periods — might not reflect the location of the decedent's lifetime health, access to health care, and state cardiovascular health promotion activities.

The proposed *Healthy People 2020* objectives for heart disease and stroke were developed to prevent premature death from cardiovascular disease by maintaining low risk for disease, controlling increased risk, detecting and treating heart attacks and strokes, and reducing disability and recurrence (*12*). Research examining health

Figure 4.3: Ground-Truth Annotations for only Row Elements

TABLE 3. Number of deaths and age-specific death rates* for coronary heart disease among adults aged ≥45 years, by age group, sex, and Hispanic ethnicity — National Vital Statistics System, United States, 2006

Column	Women						Men					
Age group (yrs)	Colu.	Column	Column	Column	Column	Column	Column	Column	Column	Column	Column	Column
	No.	Rate	(95% CI)	No.	Rate	(95% CI)	No.	Rate	(95% CI)	No.	Rate	(95% CI)
45–54	345	15.5	(13.8–17.1)	5,663	28.7	(27.9–29.4)	1,205	52.7	(49.7–55.7)	17,707	93.2	(91.8–94.6)
55–64	806	60.9	(56.7–65.1)	12,273	81.6	(80.2–83.0)	1,906	156.5	(149.5–163.6)	31,564	225.4	(222.9–227.8)
65–74	1,512	199.2	(189.2–209.2)	22,270	234.7	(231.6–237.8)	2,430	394.1	(378.5–409.8)	40,266	500.0	(495.1–504.9)
75–84	3,012	666.6	(642.8–690.4)	54,839	751.6	(745.3–757.9)	3,235	1,022.8	(987.6–1,058.1)	63,916	1,282.9	(1,273.0–1,292.9)
≥85	3,694	2,213.2	(2,141.8–2,284.5)	94,269	2,739.1	(2,721.6–2,756.6)	2,176	2,453.9	(2,350.8–2,557.0)	53,499	3,344.5	(3,316.2–3,372.9)
Total	9,369	190.0	(186.2–193.9)	189,314	344.1	(342.5–345.6)	10,952	242.0	(237.5–246.5)	206,952	434.4	(432.5–436.2)

Abbreviation: CI = confidence interval.
*Per 100,000 U.S. standard population.

TABLE 4. Number of deaths and age-specific death rates* for stroke among adults aged ≥45 years, by age group, sex, and race — National Vital Statistics System, United States, 2006

Column	Race											
Age group (yrs)	Column	Column	Column	Column	Column	Column	Column	Column	Column	Column	Column	Column
	No.	Rate	(95% CI)	No.	Rate	(95% CI)	No.	Rate	(95% CI)	No.	Rate	(95% CI)
Women												
45–54	19	—	—	109	10.4	(8.4–12.3)	875	31.3	(29.3–33.4)	1,856	10.4	(9.9–10.8)
55–64	22	16.2	(10.2–24.5)	202	28.8	(24.8–32.7)	1,090	61.1	(57.5–64.7)	3,307	24.1	(23.2–24.9)
65–74	55	78.8	(59.4–102.5)	322	80.8	(72.0–89.6)	1,565	148.9	(141.5–156.3)	6,918	79.3	(77.4–81.1)
75–84	99	267.6	(217.5–325.8)	669	284.2	(262.7–305.7)	2,701	415.6	(400.0–431.3)	21,943	321.5	(317.2–325.7)
≥85	106	648.1	(524.7–771.5)	621	777.0	(715.9–838.2)	2,901	1,060.5	(1,021.9–1,099.1)	35,698	1,102.2	(1,090.8–1,113.7)
Total	301	63.4	(56.3–70.6)	1,923	77.9	(74.5–81.4)	9,132	139.4	(136.5–142.3)	69,722	138.2	(137.1–139.2)
Men												
45–54	33	16.3	(11.2–22.9)	126	13.4	(11.1–15.8)	1,044	43.5	(40.9–46.2)	2,279	12.8	(12.3–13.4)
55–64	44	35.0	(25.4–47.0)	220	36.3	(31.5–41.1)	1,523	105.9	(100.6–111.3)	4,110	31.5	(30.5–32.4)
65–74	50	82.9	(61.5–109.3)	357	108.9	(97.6–120.2)	1,644	218.7	(208.1–229.2)	7,312	97.1	(94.9–99.3)
75–84	48	174.3	(128.5–231.1)	477	294.9	(268.5–321.4)	1,741	471.1	(449.0–493.2)	16,041	338.5	(333.2–343.7)
≥85	27	344.5	(227.0–501.2)	417	865.9	(782.8–949.0)	987	882.0	(827.0–937.0)	14,311	941.3	(925.9–956.7)
Total	202	47.6	(41.1–54.2)	1,597	76.7	(73.0–80.5)	6,939	136.9	(133.7–140.1)	44,053	98.8	(97.9–99.7)

Abbreviation: CI = confidence interval.
* Per 100,000 U.S. standard population.
† Number of deaths too small to calculate a reliable rate.

TABLE 5. Number of deaths and age-specific death rates* for stroke among adults aged ≥45 years, by age group, sex, and Hispanic ethnicity — National Vital Statistics System, United States, 2006

Column	Women						Men					
Age group (yrs)	Column	Column	Column	Column	Column	Column	Column	Colu.	Column	Column	Colu.	Column
	No.	Rate	(95% CI)	No.	Rate	(95% CI)	No.	Rate	(95% CI)	No.	Rate	(95% CI)
45–54	263	11.8	(10.4–13.2)	2,590	13.1	(12.6–13.6)	389	17.0	(15.3–18.7)	3,080	16.2	(15.6–16.8)
55–64	368	27.8	(25.0–30.7)	4,243	28.2	(27.4–29.1)	501	41.1	(37.5–44.8)	5,380	38.4	(37.4–39.4)
65–74	584	76.9	(70.7–83.2)	8,256	87.0	(85.1–88.9)	617	100.1	(92.2–108.0)	8,723	108.3	(106.0–110.6)
75–84	1,087	240.6	(226.3–254.9)	24,285	332.8	(328.6–337.0)	926	292.8	(273.9–311.6)	17,350	348.2	(343.1–353.4)
≥85	1,240	742.9	(701.6–784.3)	38,056	1,105.8	(1,094.6–1,116.9)	516	581.9	(531.7–632.1)	15,203	950.4	(935.3–965.5)
Total	3,542	71.8	(69.5–74.2)	77,430	140.7	(139.7–141.7)	2,949	65.2	(62.8–67.5)	49,736	104.4	(103.5–105.3)

Abbreviation: CI = confidence interval.
* Per 100,000 U.S. standard population.

as diabetes, which varies substantially across racial/ethnic groups. Finally, state of residence at death from CHD and stroke — diseases that often have long latency periods — might not reflect the location of the decedent's lifetime health, access to health care, and state cardiovascular health promotion activities.

The proposed *Healthy People 2020* objectives for heart disease and stroke were developed to prevent premature death from cardiovascular disease by maintaining low risk for disease, controlling increased risk, detecting and treating heart attacks and strokes, and reducing disability and recurrence (*12*). Research examining health

Figure 4.4: Ground-Truth Annotations for only Column Elements

TABLE 3. Number of deaths and age-specific death rates* for coronary heart disease among adults aged ≥45 years, by age group, sex, and Hispanic ethnicity — National Vital Statistics System, United States, 2006

Age group (yrs)	Women						Men					
	Hispanic			Non-Hispanic			Hispanic			Non-Hispanic		
	No.	Rate	(95% CI)	No.	Rate	(95% CI)	No.	Rate	(95% CI)	No.	Rate	(95% CI)
45–54	345	15.5	(13.8–17.1)	5,663	28.7	(27.9–29.4)	1,205	52.7	(49.7–55.7)	17,707	93.2	(91.8–94.6)
55–64	806	60.9	(56.7–65.1)	12,273	81.6	(80.2–83.0)	1,906	156.5	(149.5–163.6)	31,564	225.4	(222.9–227.8)
65–74	1,512	199.2	(189.2–209.2)	22,270	234.7	(231.6–237.8)	2,430	394.1	(378.5–409.8)	40,266	500.0	(495.1–504.9)
75–84	3,012	666.6	(642.8–690.4)	54,839	751.6	(745.3–757.9)	3,235	1,022.8	(987.6–1,058.1)	63,916	1,282.9	(1,273.0–1,292.9)
≥85	3,694	2,213.2	(2,141.8–2,284.5)	94,269	2,739.1	(2,721.6–2,756.6)	2,176	2,453.9	(2,350.8–2,557.0)	53,499	3,344.5	(3,316.2–3,372.9)
Total	9,369	190.0	(186.2–193.9)	189,314	344.1	(342.5–345.6)	10,952	242.0	(237.5–246.5)	206,952	434.4	(432.5–436.2)

Abbreviation: CI = confidence interval.
*Per 100,000 U.S. standard population.

TABLE 4. Number of deaths and age-specific death rates* for stroke among adults aged ≥45 years, by age group, sex, and race — National Vital Statistics System, United States, 2006

Age group (yrs)	Race											
	American Indian/Alaska Native			Asian/ Pacific Islander			Black			White		
	No.	Rate	(95% CI)	No.	Rate	(95% CI)	No.	Rate	(95% CI)	No.	Rate	(95% CI)
Women												
45–54	19	—	—	109	10.4	(8.4–12.5)	875	31.3	(29.3–33.4)	1,856	10.4	(9.9–10.8)
55–64	22	16.2	(10.2–24.5)	202	28.8	(24.8–32.7)	1,090	61.1	(57.5–64.7)	3,307	24.1	(23.2–24.9)
65–74	55	78.8	(59.4–102.5)	322	80.8	(72.0–89.6)	1,565	148.9	(141.5–156.3)	6,918	79.3	(77.4–81.1)
75–84	99	267.6	(217.5–325.8)	669	284.2	(262.7–305.7)	2,701	415.6	(400.0–431.3)	21,943	321.5	(317.2–325.7)
≥85	106	648.1	(524.7–771.5)	621	777.0	(715.9–838.2)	2,901	1,060.5	(1,021.9–1,099.1)	35,698	1,102.2	(1,090.8–1,113.7)
Total	301	63.4	(56.3–70.6)	1,923	77.9	(74.5–81.4)	9,132	139.4	(136.5–142.3)	69,722	138.2	(137.1–139.2)
Men												
45–54	33	16.3	(11.2–22.9)	126	13.4	(11.1–15.8)	1,044	43.5	(40.9–46.2)	2,279	12.8	(12.3–13.4)
55–64	44	35.0	(25.4–47.0)	220	36.3	(31.5–41.1)	1,523	105.9	(100.6–111.3)	4,110	31.5	(30.5–32.4)
65–74	50	82.9	(61.5–109.3)	357	108.9	(97.6–120.2)	1,644	218.7	(208.1–229.2)	7,312	97.1	(94.9–99.3)
75–84	48	174.3	(128.5–231.1)	477	294.9	(268.5–321.4)	1,741	471.1	(449.0–493.2)	16,041	338.5	(333.2–343.7)
≥85	27	344.5	(227.0–501.2)	417	865.9	(782.8–949.0)	987	882.0	(827.0–937.0)	14,311	941.3	(925.9–956.7)
Total	202	47.6	(41.1–54.2)	1,597	76.7	(73.0–80.5)	6,939	136.9	(133.7–140.1)	44,053	98.8	(97.9–99.7)

Abbreviation: CI = confidence interval.
* Per 100,000 U.S. standard population.
† Number of deaths too small to calculate a reliable rate.

TABLE 5. Number of deaths and age-specific death rates* for stroke among adults aged ≥45 years, by age group, sex, and Hispanic ethnicity — National Vital Statistics System, United States, 2006

Age group (yrs)	Women						Men					
	Hispanic			Non-Hispanic			Hispanic			Non-Hispanic		
	No.	Rate	(95% CI)	No.	Rate	(95% CI)	No.	Rate	(95% CI)	No.	Rate	(95% CI)
45–54	263	11.8	(10.4–13.2)	2,590	13.1	(12.6–13.6)	389	17.0	(15.3–18.7)	3,080	16.2	(15.6–16.8)
55–64	368	27.8	(25.0–30.7)	4,243	28.2	(27.4–29.1)	501	41.1	(37.5–44.8)	5,380	38.4	(37.4–39.4)
65–74	584	76.9	(70.7–83.2)	8,256	87.0	(85.1–88.9)	617	100.1	(92.2–108.0)	8,723	108.3	(106.0–110.6)
75–84	1,087	240.6	(226.3–254.9)	24,285	332.8	(328.6–337.0)	926	292.8	(273.9–311.6)	17,350	348.2	(343.1–353.4)
≥85	1,240	742.9	(701.6–784.3)	38,056	1,105.8	(1,094.6–1,116.9)	516	581.9	(531.7–632.1)	15,203	950.4	(935.3–965.5)
Total	3,542	71.8	(69.5–74.2)	77,430	140.7	(139.7–141.7)	2,949	65.2	(62.8–67.5)	49,736	104.4	(103.5–105.3)

Abbreviation: CI = confidence interval.
* Per 100,000 U.S. standard population.

as diabetes, which varies substantially across racial/ethnic groups. Finally, state of residence at death from CHD and stroke — diseases that often have long latency periods — might not reflect the location of the decedent's lifetime health, access to health care, and state cardiovascular health promotion activities.

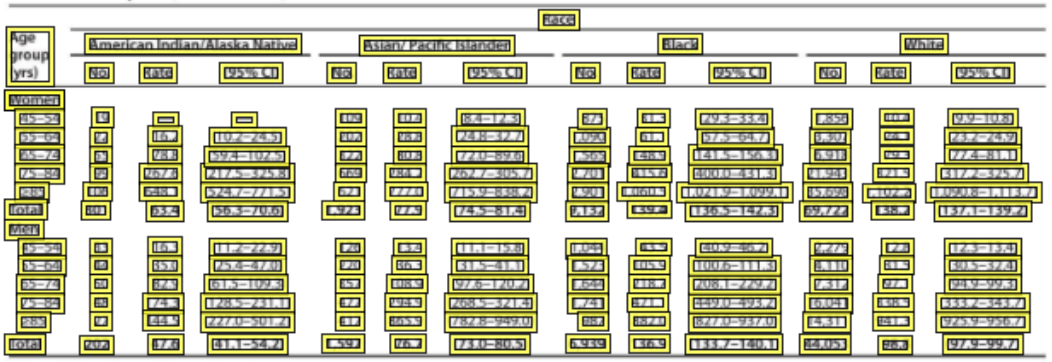
The proposed *Healthy People 2020* objectives for heart disease and stroke were developed to prevent premature death from cardiovascular disease by maintaining low risk for disease, controlling increased risk, detecting and treating heart attacks and strokes, and reducing disability and recurrence (*12*). Research examining health

Figure 4.5: Ground-Truth Annotations for Layout-Focused Structure

TABLE 3. Number of deaths and age-specific death rates* for coronary heart disease among adults aged ≥45 years, by age group, sex, and Hispanic ethnicity — National Vital Statistics System, United States, 2006

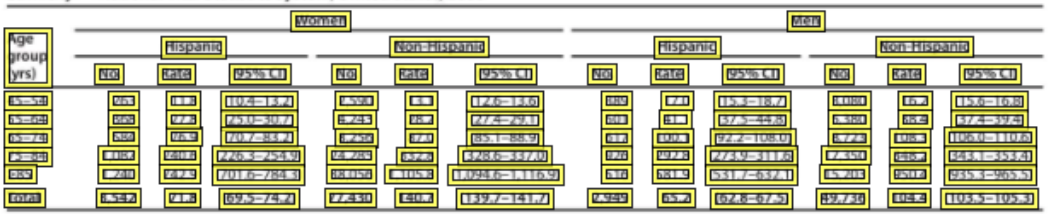
Abbreviation: CI = confidence interval.
*Per 100,000 U.S. standard population.

TABLE 4. Number of deaths and age-specific death rates* for stroke among adults aged ≥45 years, by age group, sex, and race — National Vital Statistics System, United States, 2006



Abbreviation: CI = confidence interval.
* Per 100,000 U.S. standard population.
† Number of deaths too small to calculate a reliable rate.

TABLE 5. Number of deaths and age-specific death rates* for stroke among adults aged ≥45 years, by age group, sex, and Hispanic ethnicity — National Vital Statistics System, United States, 2006



Abbreviation: CI = confidence interval.
* Per 100,000 U.S. standard population.

as diabetes, which varies substantially across racial/ethnic groups. Finally, state of residence at death from CHD and stroke — diseases that often have long latency periods — might not reflect the location of the decedent's lifetime health, access to health care, and state cardiovascular health promotion activities.

The proposed *Healthy People 2020* objectives for heart disease and stroke were developed to prevent premature death from cardiovascular disease by maintaining low risk for disease, controlling increased risk, detecting and treating heart attacks and strokes, and reducing disability and recurrence (*12*). Research examining health

Figure 4.6: Ground-Truth Annotations for Content-Focused Structure

4.1.2 Table Structure Detection in Cell Format

Tables are made up of cells, which are the most fundamental components. Defining tables in the cell format reduce the complexity of the structure detection problem. Spanning cells more than one row or column can be ground-truthed easily. The overlapping problem is solved by reducing the table elements to their simplest component. Also, there is no need for separate models that specialize in different objects other than cell objects. In the annotations of cell format, we followed two different strategies: 1) Layout-focused cell structure, 2) Content-focused cell structure. Layout-focused annotations are done by following the ruling lines. When the table does not have ruling lines, the center point of the separating regions is considered as the boundary. In content-focused annotations, each cell content is surrounded closely by the bounding box regardless of the presence of the ruling lines. The sample layout-focused annotations and content-focused annotations are displayed in Figure 4.5 and Figure 4.6 respectively.

4.2 Annotation Bootstrapping Strategy

Deep learning-based solutions are data-driven and require domain-specific labelled data. For the task of table structure detection, publicly available annotated data of good quality is scarce. There have been released publicly available large-scale datasets in recent years; however, they differ in style due to the various definitions of table structure as well as domain. The challenges of annotating data, especially for the table structure detection task, are thoroughly covered in the preceding section. We investigated the best structure definition to simplify the problem and reduce the ambiguities to the minimum. We propose annotation bootstrapping to make the labelling process more efficient and less error-prone.

Deep learning models usually take document images as input since it is the most generic format. However, many PDF-based approaches exist in the literature, and some of them are explained in Chapter 2. Utilizing a robust pdf-based table structure detection tool to bootstrap annotations can significantly reduce the amount of time for the labelling process. Another benefit of this strategy is that it allows discovering where these tools perform poorly and what improvements they need. We utilized Pdfplumber [16] to obtain content-focused annotations of table structures. Pdfplumber has two strategies to detect tables. First, it exploits only vertical and horizontal line information to detect tables. Second, text strategy where visual clues such as the vertical and horizontal alignment of the text, spaces between separator regions, and line information are used to detect tables. We combine both approaches and use the one that recognizes a higher number of tables. If

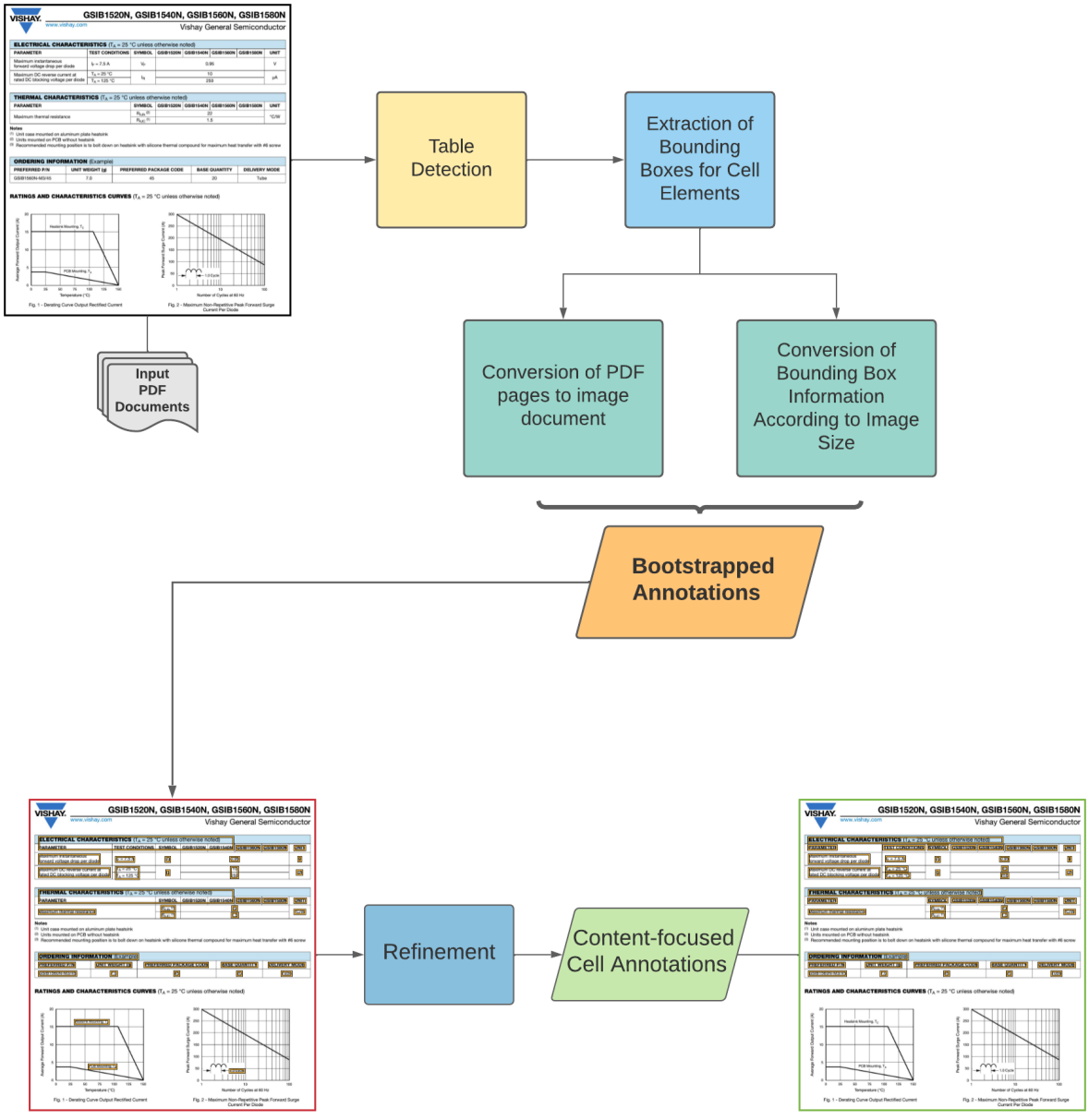


Figure 4.7: Pipeline of the Annotation Bootstrapping Strategy

the number of detected tables is equal, line strategy is chosen since its detections are more accurate than text strategy. Following the table detection, locations of word elements are extracted based on the horizontal and vertical spacing threshold. Extracted coordinates are converted according to the size of the images. The resulting annotated data is reprocessed

to get its final form. This process is illustrated in Figure 4.7. This method revealed that determining the boundaries for multi-line cell elements or spanning rows/columns is challenging for Pdfplumber.

4.3 Dataset and Training Details

ICDAR 2013 [68] is a benchmark dataset in table detection and table structure detection tasks and has been widely used. There are 128 image documents that each image has at least one table. It is a small dataset; however, an excellent choice to conduct experiments for table structure definition. We have confined our research in this section to the ICDAR 2013 dataset because labelling it in all table structure formats is tedious and time-consuming. Dataset is randomly divided into training and test sets with a ratio of 80/20. Images in the dataset are hand labelled with each type of format that is explained in Section 4.1. All models are implemented by using the detectron2 framework [69]. The Mask R-CNN with ResNet101 backbone is chosen because it is less complex yet provides promising performance on instance segmentation and table structure detection tasks. The details of the model are explained in Chapter 3. Models are trained on the Mist GPU cluster that has 4 Tesla V 100 GPUs with 32GB VRAM per node. All models are trained for 500 epochs on a single GPU. Images are fed into the model randomly at each epoch. Learning rate and batch size are defined as 0.002 and 2, respectively. The training time is around 8 hours for the model with these configurations. We refer readers to [70] to set hyperparameters for different GPU configurations. However, unlike general object detection tasks, table structure detection task requires lower learning rates due to very high amount of object per image. The training loss and accuracy curves were examined to ensure that each model was fully trained.

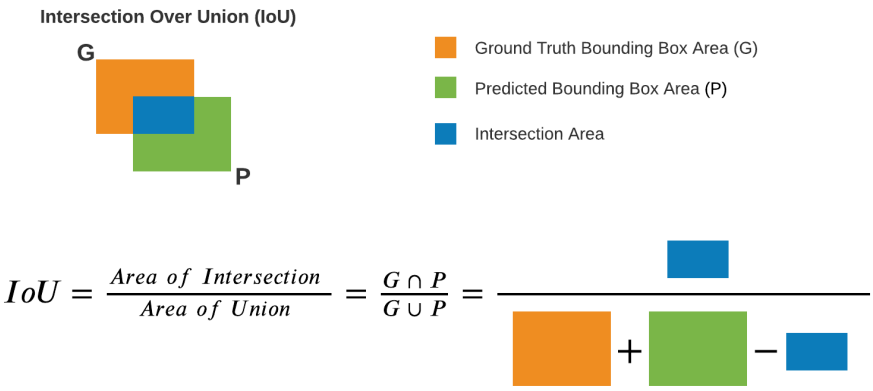


Figure 4.8: Intersection over Union (IoU) concept

4.4 Evaluation of Table Structure Detection Results Under Different Formats

As explained in the previous section, table structures are defined under three different formats: 1) row/column, 2) layout-focused cell, and 3) content-focused cell. In this section, the performance of our base deep learning model will be investigated in detail with numerical and visual data on structure detection task. Table structure detection is localizing table structure elements that are defined in this chapter. It is worth to mention that it does not include creating relations between table entities. Table 4.1 gives the precision, recall and F1-score. The concept of Intersection over Union (IoU) is used to determine true detections and false detections. IoU is the ratio of overlapping area to the combined area of ground truth and predicted bounding box. Evaluation metrics are calculated for 50% IoU threshold and over with 5% incremental steps until 95%. The mean values were derived by averaging the values between 50% and 95%. If the IoU value is under 50%, detections are considered as false positive. IoU concept is illustrated in Fig. 4.8. An IoU threshold of 50% means half of the predicted area overlaps with the ground-truth area, and the object is loosely localized, where 100% IoU means the object is perfectly localized. Since we focus on the locating table elements, the common metric for object detection tasks IoU is used for evaluation, which is widely adopted by the research community.

The Structure detection model provided 66.2% average precision, 71.3% average recall and 68.7% average F1-score when trained to detect both rows and columns simultaneously.

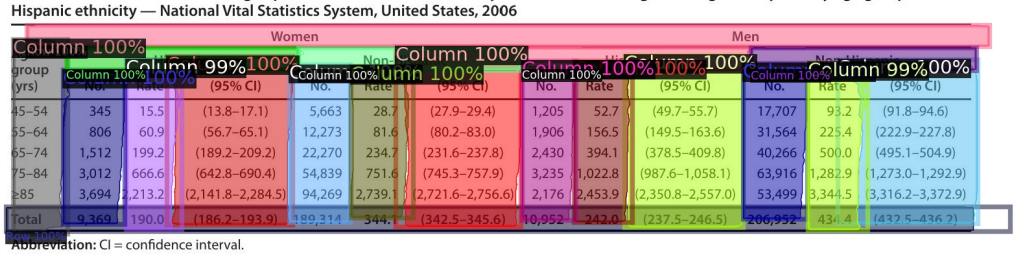
Table 4.1: Table Structure Detection Results Under Different Formats

			IoU Threshold				
Metric	Structure Format	Average	50%	60%	75%	80%	90%
Precision	Row/Column (Combined)	66.2	84.4	82.4	74.7	68.0	38.9
	Row	75.6	92.9	92.5	83.4	80.1	50.9
	Column	75.9	92.6	89.0	82.8	76.9	53.6
	Layout-Focused Cell	56.4	93.3	86.9	60.0	40.7	11.4
	Content-Focused Cell	70.5	93.3	93.3	86.3	77.7	22.6
Recall	Row/Column (Combined)	71.3	85.8	84.5	78.3	73.8	49.9
	Row	81.7	94.0	93.3	88.4	86.2	65.8
	Column	82.5	96.9	94.2	88.1	83.6	63.7
	Layout-Focused Cell	64.8	95.0	90.8	71.6	56.5	25.0
	Content-Focused Cell	77.4	94.5	94.4	90.2	85.5	45.1
F-1 Score	Row/Column (Combined)	68.7	85.1	83.4	76.5	70.8	43.7
	Row	78.5	93.4	92.9	85.8	83.0	57.4
	Column	79.1	94.7	91.5	85.4	80.1	58.2
	Layout-Focused Cell	60.3	94.1	88.8	65.3	47.3	15.7
	Content-Focused Cell	73.8	93.9	93.8	88.2	81.4	30.1

Although it provides the lower performance in average values than separate models, the model performs promising results at the low-level IoU threshold and performs better over 80% IoU thresholds. An example of inference results for row and column detection is displayed in Figure 4.9. Given the different characteristics of row and column objects and their highly overlapping nature, we trained separate models that specialized on rows and columns, respectively. Both row and column detection models provided the best average precision, recall and F1-score, and values almost identical. Models performed flawlessly at lower IoU thresholds by obtaining over 93.4% and 94.7% F1-score and provided very promising results with F1-score of 57.4% and 58.2% at 90 percent IoU threshold. Inference results for separate models are presented in Figure 4.10 and 4.11. As the numerical results suggest, it can be clearly seen that separate models provide smoother detections than the combined model.

Cell detection models are proposed to handle better spanning rows and columns. The layout-focused cell detection model seems to achieve the lowest average results in all metrics. However, it performs similarly to separate row/column models at lower IoU thresholds and achieves the highest F1-score following the separate models. The layout-focused model

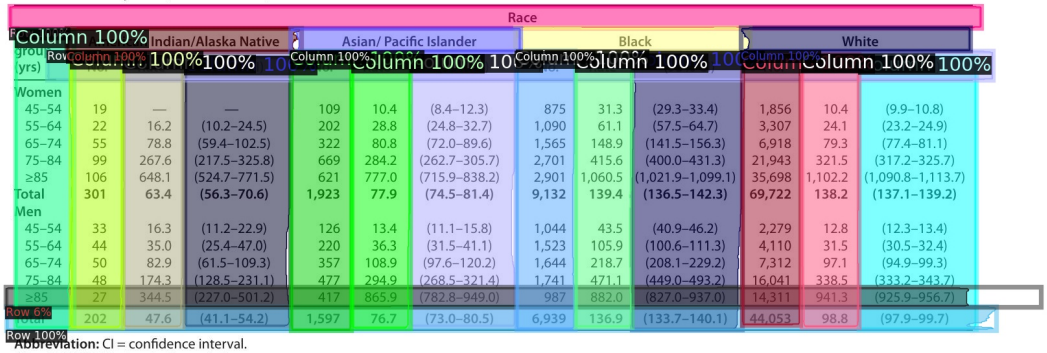
TABLE 3. Number of deaths and age-specific death rates* for coronary heart disease among adults aged ≥45 years, by age group, sex, and Hispanic ethnicity — National Vital Statistics System, United States, 2006



age group (yrs)	Women						Men					
	Hispanic			Non-Hispanic			Hispanic			Non-Hispanic		
	No.	Rate	(95% CI)	No.	Rate	(95% CI)	No.	Rate	(95% CI)	No.	Rate	(95% CI)
45–54	345	15.5	(13.8–17.1)	5,663	28.7	(27.9–29.4)	1,205	52.7	(49.7–55.7)	17,707	93.2	(91.8–94.6)
55–64	806	60.9	(56.7–65.1)	12,273	81.6	(80.2–83.0)	1,906	156.5	(149.5–163.6)	31,564	225.4	(222.9–227.8)
65–74	1,512	199.2	(189.2–209.2)	22,270	234.7	(231.6–237.8)	2,430	394.1	(378.5–409.8)	40,266	500.0	(495.1–504.9)
75–84	3,012	666.6	(642.8–690.4)	54,839	751.6	(745.3–757.9)	3,235	1,022.8	(987.6–1,058.1)	63,916	1,282.9	(1,273.0–1,292.9)
≥85	3,694	1,213.2	(2,141.8–2,284.5)	94,269	2,739.1	(2,721.6–2,756.6)	2,176	2,453.9	(2,350.8–2,557.0)	53,499	3,344.5	(3,316.2–3,372.9)
Total	9,369	190.0	(186.2–193.9)	189,314	344.1	(342.5–345.6)	10,952	242.0	(237.5–246.5)	206,952	434.4	(432.5–436.2)

Abbreviation: CI = confidence interval.
*Per 100,000 U.S. standard population.

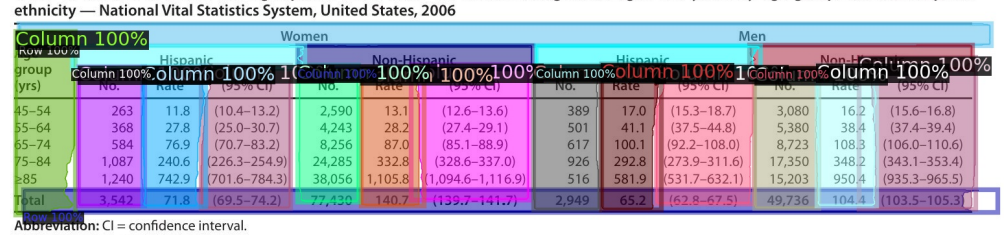
TABLE 4. Number of deaths and age-specific death rates* for stroke among adults aged ≥45 years, by age group, sex, and race — National Vital Statistics System, United States, 2006



age group (yrs)	Race											
	American Indian/Alaska Native			Asian/ Pacific Islander			Black			White		
	No.	Rate	(95% CI)	No.	Rate	(95% CI)	No.	Rate	(95% CI)	No.	Rate	(95% CI)
Women												
45–54	19	—	—	109	10.4	(8.4–12.3)	875	31.3	(29.3–33.4)	1,856	10.4	(9.9–10.8)
55–64	22	16.2	(10.2–24.5)	202	28.8	(24.8–32.7)	1,090	61.1	(57.5–64.7)	3,307	24.1	(23.2–24.9)
65–74	55	78.8	(59.4–102.5)	322	80.8	(72.0–89.6)	1,565	148.9	(141.5–156.3)	6,918	79.3	(77.4–81.1)
75–84	99	267.6	(217.5–325.8)	669	284.2	(262.7–305.7)	2,701	415.6	(400.0–431.3)	21,943	321.5	(317.2–325.7)
≥85	106	648.1	(524.7–771.5)	621	777.0	(715.9–838.2)	2,901	1,060.5	(1,021.9–1,099.1)	35,698	1,102.2	(1,090.8–1,113.7)
Total	301	63.4	**(56.3–70.6)**	1,923	77.9	**(74.5–81.4)**	9,132	139.4	**(136.5–142.3)**	69,722	138.2	**(137.1–139.2)**
Men												
45–54	33	16.3	(11.2–22.9)	126	13.4	(11.1–15.8)	1,044	43.5	(40.9–46.2)	2,279	12.8	(12.3–13.4)
55–64	44	35.0	(25.4–47.0)	220	36.3	(31.5–41.1)	1,523	105.9	(100.6–111.3)	4,110	31.5	(30.5–32.4)
65–74	50	82.9	(61.5–109.3)	357	108.9	(97.6–120.2)	1,644	218.7	(208.1–229.2)	7,312	97.1	(94.9–99.3)
75–84	48	174.3	(128.5–231.1)	477	294.9	(268.5–321.4)	1,741	471.1	(449.0–493.2)	16,041	338.5	(333.2–343.7)
≥85	27	344.5	(227.0–501.2)	417	865.9	(782.8–949.0)	987	882.0	(827.0–937.0)	14,311	941.3	(925.9–956.7)
Total	202	47.6	(41.1–54.2)	1,597	76.7	(73.0–80.5)	6,939	136.9	(133.7–140.1)	44,053	98.8	(97.9–99.7)

Abbreviation: CI = confidence interval.
* Per 100,000 U.S. standard population.
† Number of deaths too small to calculate a reliable rate.

TABLE 5. Number of deaths and age-specific death rates* for stroke among adults aged ≥45 years, by age group, sex, and Hispanic ethnicity — National Vital Statistics System, United States, 2006



age group (yrs)	Women						Men					
	Hispanic			Non-Hispanic			Hispanic			Non-Hispanic		
	No.	Rate	(95% CI)	No.	Rate	(95% CI)	No.	Rate	(95% CI)	No.	Rate	(95% CI)
45–54	263	11.8	(10.4–13.2)	2,590	13.1	(12.6–13.6)	389	17.0	(15.3–18.7)	3,080	16.2	(15.6–16.8)
55–64	368	27.8	(25.0–30.7)	4,243	28.2	(27.4–29.1)	501	41.1	(37.5–44.8)	5,380	38.4	(37.4–39.4)
65–74	584	76.9	(70.7–83.2)	8,256	87.0	(85.1–88.9)	617	100.1	(92.2–108.0)	8,723	108.3	(106.0–110.6)
75–84	1,087	240.6	(226.3–254.9)	24,285	332.8	(328.6–337.0)	926	292.8	(273.9–311.6)	17,350	348.2	(343.1–353.4)
≥85	1,240	742.9	(701.6–784.3)	38,056	1,105.8	(1,094.6–1,116.9)	516	581.9	(531.7–632.1)	15,203	950.4	(935.3–965.5)
Total	3,542	71.8	(69.5–74.2)	77,430	140.7	(139.7–141.7)	2,949	65.2	(62.8–67.5)	49,736	104.4	(103.5–105.3)

Abbreviation: CI = confidence interval.
* Per 100,000 U.S. standard population.

as diabetes, which varies substantially across racial/ethnic groups. Finally, state of residence at death from CHD and stroke — diseases that often have long latency periods — might not reflect the location of the decedent's lifetime health, access to health care, and state cardiovascular health promotion activities.

The proposed *Healthy People 2020* objectives for heart disease and stroke were developed to prevent premature death from cardiovascular disease by maintaining low risk for disease, controlling increased risk, detecting and treating heart attacks and strokes, and reducing disability and recurrence (*12*). Research examining health

Figure 4.9: Inference Results with Row and Column Detection (Combined)

TABLE 3. Number of deaths and age-specific death rates* for coronary heart disease among adults aged ≥45 years, by age group, sex, and Hispanic ethnicity — National Vital Statistics System, United States, 2006

	Women						Men					
	Hispanic			Non-Hispanic			Hispanic			Non-Hispanic		
Age group (yrs)	No.	Rate	(95% CI)	No.	Rate	(95% CI)	No.	Rate	(95% CI)	No.	Rate	(95% CI)
45–54	345	15.5	(13.8–17.1)	5,663	28.7	(27.9–29.4)	1,205	52.7	(49.7–55.7)	17,707	93.2	(91.8–94.6)
55–64	806	60.9	(56.7–65.1)	12,273	81.6	(80.2–83.0)	1,906	156.5	(149.5–163.6)	31,564	225.4	(222.9–227.8)
65–74	1,512	199.2	(189.2–209.2)	22,270	234.7	(231.6–237.8)	2,430	394.1	(378.5–409.8)	40,266	500.0	(495.1–504.9)
75–84	3,012	666.6	(642.8–690.4)	54,839	751.6	(745.3–757.9)	3,235	1,022.8	(987.6–1,058.1)	63,916	1,282.9	(1,273.0–1,292.9)
≥85	3,694	2,213.2	(2,141.8–2,284.5)	94,269	2,739.1	(2,721.6–2,756.6)	2,176	2,453.9	(2,350.8–2,557.0)	53,499	3,344.5	(3,316.2–3,372.9)
Total	9,369	190.0	(186.2–193.9)	189,314	344.1	(342.5–345.6)	10,952	242.0	(237.5–246.5)	206,952	434.4	(432.5–436.2)

Abbreviation: CI = confidence interval.
*Per 100,000 U.S. standard population.

TABLE 4. Number of deaths and age-specific death rates* for stroke among adults aged ≥45 years, by age group, sex, and race — National Vital Statistics System, United States, 2006

	Race											
	American Indian/Alaska Native			Asian/ Pacific Islander			Black			White		
Age group (yrs)	No.	Rate	(95% CI)	No.	Rate	(95% CI)	No.	Rate	(95% CI)	No.	Rate	(95% CI)
Women												
45–54	19	—	—	109	10.4	(8.4–12.3)	875	31.3	(29.3–33.4)	1,856	10.4	(9.9–10.8)
55–64	22	16.2	(10.2–24.5)	202	28.8	(24.8–32.7)	1,090	61.1	(57.5–64.7)	3,307	24.1	(23.2–24.9)
65–74	55	78.8	(59.4–102.5)	322	80.8	(72.0–89.6)	1,565	148.9	(141.5–156.3)	6,918	79.3	(77.4–81.1)
75–84	99	267.6	(217.5–325.8)	669	284.2	(262.7–305.7)	2,701	415.6	(400.0–431.3)	21,943	321.5	(317.2–325.7)
≥85	106	648.1	(524.7–771.5)	621	777.0	(715.9–838.2)	2,901	1,060.5	(1,021.9–1,099.1)	35,698	1,102.2	(1,090.8–1,113.7)
Total	301	63.4	(56.3–70.6)	1,923	77.9	(74.5–81.4)	9,132	139.4	(136.5–142.3)	69,722	138.2	(137.1–139.2)
Men												
45–54	33	16.3	(11.2–22.9)	126	13.4	(11.1–15.8)	1,044	43.5	(40.9–46.2)	2,279	12.8	(12.3–13.4)
55–64	44	35.0	(25.4–47.0)	220	36.3	(31.5–41.1)	1,523	105.9	(100.6–111.3)	4,110	31.5	(30.5–32.4)
65–74	50	82.9	(61.5–109.3)	357	108.9	(97.6–120.2)	1,644	218.7	(208.1–229.2)	7,312	97.1	(94.9–99.3)
75–84	48	174.3	(128.5–231.1)	477	294.9	(268.5–321.4)	1,741	471.1	(449.0–493.2)	16,041	338.5	(333.2–343.7)
≥85	27	344.5	(227.0–501.2)	417	865.9	(782.8–949.0)	987	882.0	(827.0–937.0)	14,311	941.3	(925.9–956.7)
Total	202	47.6	(41.1–54.2)	1,597	76.7	(73.0–80.5)	6,939	136.9	(133.7–140.1)	44,053	98.8	(97.9–99.7)

Abbreviation: CI = confidence interval.
* Per 100,000 U.S. standard population.
† Number of deaths too small to calculate a reliable rate.

TABLE 5. Number of deaths and age-specific death rates* for stroke among adults aged ≥45 years, by age group, sex, and Hispanic ethnicity — National Vital Statistics System, United States, 2006

	Women						Men					
	Hispanic			Non-Hispanic			Hispanic			Non-Hispanic		
Age group (yrs)	No.	Rate	(95% CI)	No.	Rate	(95% CI)	No.	Rate	(95% CI)	No.	Rate	(95% CI)
45–54	263	11.8	(10.4–13.2)	2,590	13.1	(12.6–13.6)	389	17.0	(15.3–18.7)	3,080	16.2	(15.6–16.8)
55–64	368	27.8	(25.0–30.7)	4,243	28.2	(27.4–29.1)	501	41.1	(37.5–44.8)	5,380	38.4	(37.4–39.4)
65–74	584	76.9	(70.7–83.2)	8,256	87.0	(85.1–88.9)	617	100.1	(92.2–108.0)	8,723	108.3	(106.0–110.6)
75–84	1,087	240.6	(226.3–254.9)	24,285	332.8	(328.6–337.0)	926	292.8	(273.9–311.6)	17,350	348.2	(343.1–353.4)
≥85	1,240	742.9	(701.6–784.3)	38,056	1,105.8	(1,094.6–1,116.9)	516	581.9	(531.7–632.1)	15,203	950.4	(935.3–965.5)
Total	3,542	71.8	(69.5–74.2)	77,430	140.7	(139.7–141.7)	2,949	65.2	(62.8–67.5)	49,736	104.4	(103.5–105.3)

Abbreviation: CI = confidence interval.
* Per 100,000 U.S. standard population.

as diabetes, which varies substantially across racial/ethnic groups. Finally, state of residence at death from CHD and stroke — diseases that often have long latency periods — might not reflect the location of the decedent's lifetime health, access to health care, and state cardiovascular health promotion activities.

The proposed *Healthy People 2020* objectives for heart disease and stroke were developed to prevent premature death from cardiovascular disease by maintaining low risk for disease, controlling increased risk, detecting and treating heart attacks and strokes, and reducing disability and recurrence (*12*). Research examining health

Figure 4.10: Inference Results with Only Row Detection

TABLE 3. Number of deaths and age-specific death rates* for coronary heart disease among adults aged ≥45 years, by age group, sex, and Hispanic ethnicity — National Vital Statistics System, United States, 2006

Age group (yrs)	Women						Men					
	Hispanic			Non-Hispanic			Hispanic			Non-Hispanic		
	No.	Rate	95% CI	No.	Rate	95% CI	No.	Rate	95% CI	No.	Rate	95% CI
45–54	345	15.5	(13.8–17.1)	5,663	28.7	(27.9–29.4)	1,205	52.7	(49.7–55.7)	17,707	93.2	(91.8–94.6)
55–64	806	60.9	(56.7–65.1)	12,273	81.6	(80.2–83.0)	1,906	156.5	(149.5–163.6)	31,564	225.4	(222.9–227.8)
65–74	1,512	199.2	(189.2–209.2)	22,270	234.7	(231.6–237.8)	2,430	394.1	(378.5–409.8)	40,266	500.0	(495.1–504.9)
75–84	3,012	666.6	(642.8–690.4)	54,839	751.6	(745.3–757.9)	3,235	1,022.8	(987.6–1,058.1)	63,916	1,282.9	(1,273.0–1,292.9)
≥85	3,694	2,213.2	(2,141.8–2,284.5)	94,269	2,739.1	(2,721.6–2,756.6)	2,176	2,453.9	(2,350.8–2,557.0)	53,499	3,344.5	(3,316.2–3,372.9)
Total	9,369	190.0	(186.2–193.9)	189,314	344.1	(342.5–345.6)	10,952	242.0	(237.5–246.5)	206,952	434.4	(432.5–436.2)

Abbreviation: CI = confidence interval.
*Per 100,000 U.S. standard population.

TABLE 4. Number of deaths and age-specific death rates* for stroke among adults aged ≥45 years, by age group, sex, and race — National Vital Statistics System, United States, 2006

Age group (yrs)	Race											
	American Indian/Alaska Native			Asian/ Pacific Islander			Black			White		
	No.	Rate	95% CI	No.	Rate	95% CI	No.	Rate	95% CI	No.	Rate	95% CI
Women												
45–54	19	—	—	109	10.4	(8.4–12.3)	875	31.3	(29.3–33.4)	1,856	10.4	(9.9–10.8)
55–64	22	16.2	(10.2–24.5)	202	28.8	(24.8–32.7)	1,090	61.1	(57.5–64.7)	3,307	24.1	(23.2–24.9)
65–74	55	78.8	(59.4–102.5)	322	80.8	(72.0–89.6)	1,565	148.9	(141.5–156.3)	6,918	79.3	(77.4–81.1)
75–84	99	267.6	(217.5–325.8)	669	284.2	(262.7–305.7)	2,701	415.6	(400.0–431.3)	21,943	321.5	(317.2–325.7)
≥85	106	648.1	(524.7–771.5)	621	777.0	(715.9–838.2)	2,901	1,060.5	(1,021.9–1,099.1)	35,698	1,102.2	(1,090.8–1,113.7)
Total	**301**	**63.4**	**(56.3–70.6)**	**1,923**	**77.9**	**(74.5–81.4)**	**9,132**	**139.4**	**(136.5–142.3)**	**69,722**	**138.2**	**(137.1–139.2)**
Men												
45–54	33	16.3	(11.2–22.9)	126	13.4	(11.1–15.8)	1,044	43.5	(40.9–46.2)	2,279	12.8	(12.3–13.4)
55–64	44	35.0	(25.4–47.0)	220	36.3	(31.5–41.1)	1,523	105.9	(100.6–111.3)	4,110	31.5	(30.5–32.4)
65–74	50	82.9	(61.5–109.3)	357	108.9	(97.6–120.2)	1,644	218.7	(208.1–229.2)	7,312	97.1	(94.9–99.3)
75–84	48	174.3	(128.5–231.1)	477	294.9	(268.5–321.4)	1,741	471.1	(449.0–493.2)	16,041	338.5	(333.2–343.7)
≥85	27	344.5	(227.0–501.2)	417	865.9	(782.8–949.0)	987	882.0	(827.0–937.0)	14,311	941.3	(925.9–956.7)
Total	**202**	**47.6**	**(41.1–54.2)**	**1,597**	**76.7**	**(73.0–80.5)**	**6,939**	**136.9**	**(133.7–140.1)**	**44,053**	**98.8**	**(97.9–99.7)**

Abbreviation: CI = confidence interval.
* Per 100,000 U.S. standard population.
† Number of deaths too small to calculate a reliable rate.

TABLE 5. Number of deaths and age-specific death rates* for stroke among adults aged ≥45 years, by age group, sex, and Hispanic ethnicity — National Vital Statistics System, United States, 2006

Age group (yrs)	Women						Men					
	Hispanic			Non-Hispanic			Hispanic			Non-Hispanic		
	No.	Rate	95% CI	No.	Rate	95% CI	No.	Rate	95% CI	No.	Rate	95% CI
45–54	263	11.8	(10.4–13.2)	2,590	13.1	(12.6–13.6)	389	17.0	(15.3–18.7)	3,080	16.2	(15.6–16.8)
55–64	368	27.8	(25.0–30.7)	4,243	28.2	(27.4–29.1)	501	41.1	(37.5–44.8)	5,380	38.4	(37.4–39.4)
65–74	584	76.9	(70.7–83.2)	8,256	87.0	(85.1–88.9)	617	100.1	(92.2–108.0)	8,723	108.3	(106.0–110.6)
75–84	1,087	240.6	(226.3–254.9)	24,285	332.8	(328.6–337.0)	926	292.8	(273.9–311.6)	17,350	348.2	(343.1–353.4)
≥85	1,240	742.9	(701.6–784.3)	38,056	1,105.8	(1,094.6–1,116.9)	516	581.9	(531.7–632.1)	15,203	950.4	(935.3–965.5)
Total	3,542	71.8	(69.5–74.2)	77,430	140.7	(139.7–141.7)	2,949	65.2	(62.8–67.5)	49,736	104.4	(103.5–105.3)

Abbreviation: CI = confidence interval.
* Per 100,000 U.S. standard population.

as diabetes, which varies substantially across racial/ethnic groups. Finally, state of residence at death from CHD and stroke — diseases that often have long latency periods — might not reflect the location of the decedent's lifetime health, access to health care, and state cardiovascular health promotion activities.

The proposed *Healthy People 2020* objectives for heart disease and stroke were developed to prevent premature death from cardiovascular disease by maintaining low risk for disease, controlling increased risk, detecting and treating heart attacks and strokes, and reducing disability and recurrence (*12*). Research examining health

Figure 4.11: Inference Results with Only Column Detection

TABLE 3. Number of deaths and age-specific death rates* for coronary heart disease among adults aged ≥45 years, by age group, sex, and Hispanic ethnicity — National Vital Statistics System, United States, 2006

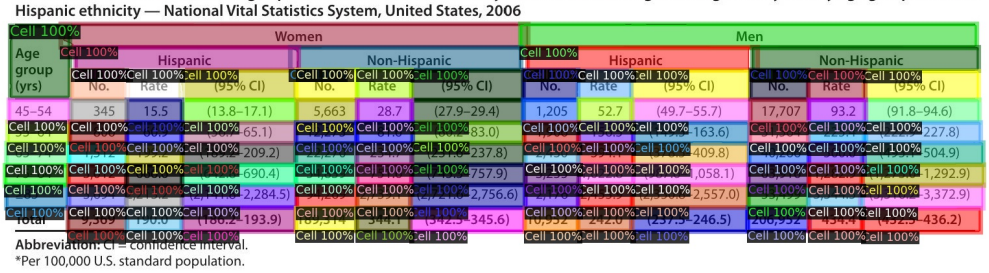


TABLE 4. Number of deaths and age-specific death rates* for stroke among adults aged ≥45 years, by age group, sex, and race — National Vital Statistics System, United States, 2006

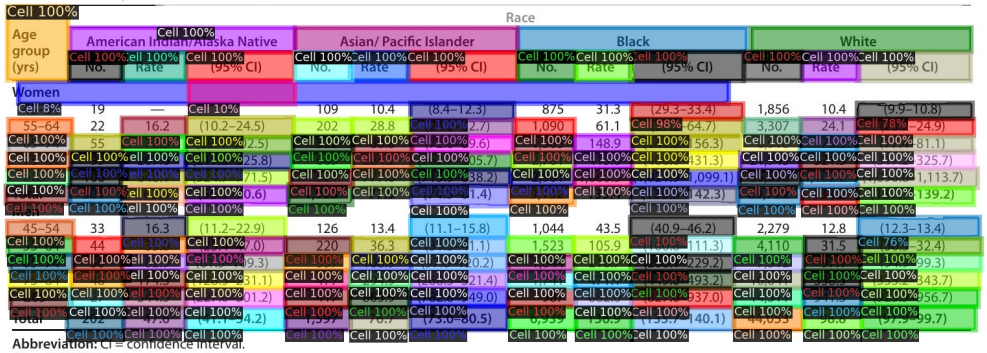
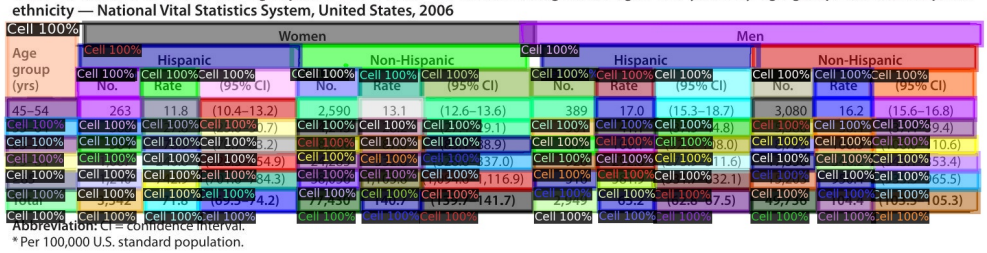


TABLE 5. Number of deaths and age-specific death rates* for stroke among adults aged ≥45 years, by age group, sex, and Hispanic ethnicity — National Vital Statistics System, United States, 2006



as diabetes, which varies substantially across racial/ethnic groups. Finally, state of residence at death from CHD and stroke — diseases that often have long latency periods — might not reflect the location of the decedent's lifetime health, access to health care, and state cardiovascular health promotion activities.

The proposed *Healthy People 2020* objectives for heart disease and stroke were developed to prevent premature death from cardiovascular disease by maintaining low risk for disease, controlling increased risk, detecting and treating heart attacks and strokes, and reducing disability and recurrence (*12*). Research examining health

Figure 4.12: Inference Results with Layout-Focused Cell Detection

TABLE 3. Number of deaths and age-specific death rates* for coronary heart disease among adults aged ≥45 years, by age group, sex, and Hispanic ethnicity — National Vital Statistics System, United States, 2006

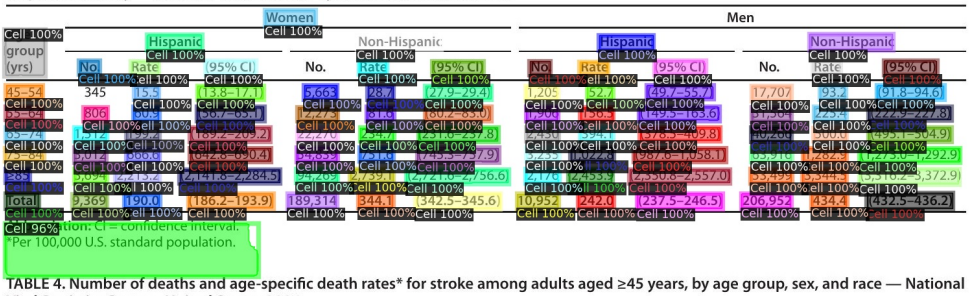


TABLE 4. Number of deaths and age-specific death rates* for stroke among adults aged ≥45 years, by age group, sex, and race — National Vital Statistics System, United States, 2006

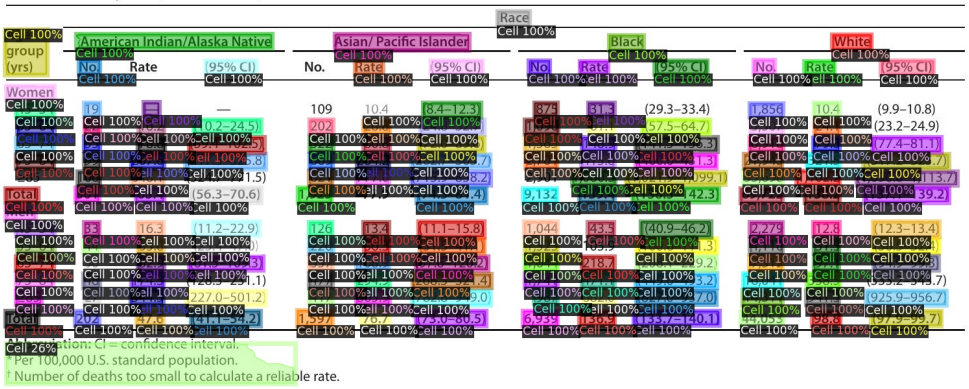
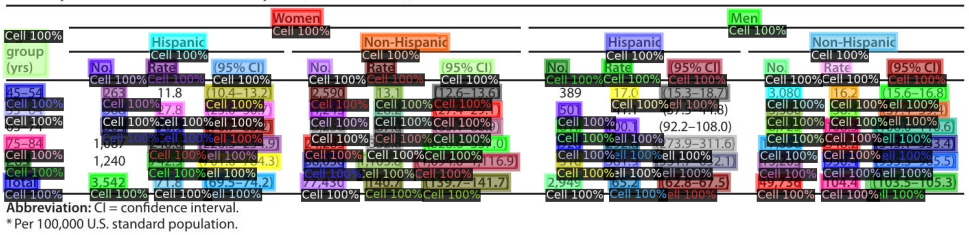


TABLE 5. Number of deaths and age-specific death rates* for stroke among adults aged ≥45 years, by age group, sex, and Hispanic ethnicity — National Vital Statistics System, United States, 2006



as diabetes, which varies substantially across racial/ethnic groups. Finally, state of residence at death from CHD and stroke — diseases that often have long latency periods — might not reflect the location of the decedent's lifetime health, access to health care, and state cardiovascular health promotion activities.

The proposed *Healthy People 2020* objectives for heart disease and stroke were developed to prevent premature death from cardiovascular disease by maintaining low risk for disease, controlling increased risk, detecting and treating heart attacks and strokes, and reducing disability and recurrence (*12*). Research examining health

Figure 4.13: Inference Results with Content-Focused Detection

performs poorly at higher IoU thresholds that cause the lower average results. There are two reasons behind the poor performance at high IoU thresholds. First, cell objects are smaller than row and column objects. Second, cell objects appear densely on image documents since the number of elements drastically increases in the cell format. The content-focused model achieves the best results after separate row and column models with an F1-score of 73.8%. Unlike the layout-focused model, the content-focused model performs similarly to or even better than separate row/column models until an IoU threshold of 80%. Its performance, however, drops sharply beyond the 80% threshold yet still doubles the performance of the layout-focused model. Inference results of the layout-focused model and the content-focused model are exemplified in Figure 4.12 and 4.13. respectively.

Overall, the content-focused model and separate row and column models are the top-performing models. The poor performance of the content-focused model at the 90% IoU threshold and beyond causes around 5% performance difference in the average results. However, realizing 100 percent overlapping between the ground truth and prediction is not trivial and achievable even for general object detection tasks. Considering the existence of a high number of objects in document images when tables are defined in cell format, it is acceptable to have lower results at higher Iou thresholds. Moreover, the content-focused model is more advantageous thanks to its flexibility to tackle spanning rows and columns problems. For these reasons, we propose the content-focused cell detection model for table structure detection.

4.5 Summary

In this chapter, ground-truthing table structures in various formats is investigated for the table structure detection task. The most optimal approach in terms of performance and problem simplicity is offered as content-focused cell detection. This methodology achieves an F1-score of 93.9% at 50% of the IoU threshold and 73.8% on average. The ambiguity of detecting table elements without ruling lines and the spanning row/column problem are significantly handled. Additionally, an annotation bootstrapping strategy is proposed to obtain high-quality labelled data in an efficient way for table structure detection task.

Chapter 5

Structure Detection under Cropped versus Uncropped Tabular Images

This chapter presents the cropped and uncropped methods to perform table structure detection under two sets of experiments. It is worth to note that developing a network architecture is not the focus of work in this chapter; however, the architectures of already implemented networks are presented in chapter 3.

5.1 The Cropped and Uncropped Approaches

Extraction of table contents involve two consecutive stages: 1) table detection, 2) table structure recognition. Table detection is the localization of the table object on the image layout. Table structure recognition task includes localizing table structure elements (Structure Detection), e.g., row, column or cell and detecting relationships between these elements. It is assumed that table location is perfectly determined by a table detection model to perform structure detection task in state of the art. We adopt this approach by cropping tables from document images to perform structure detection task. This process is illustrated in Fig. 5.1. On the other hand, document images have plots, figures or textual areas surrounding tables. These non-table objects either may help the model to generalize better on table structure or degrade the performance. These document images are referred to as uncropped or regular, and the structure detection process for this kind of image is summarized in Fig. 5.1. We conduct two sets of experiments to compare the impact of cropped and uncropped sets on table structure detection performance. First, only ICDAR

2017 dataset is used for the comparison of cropped and uncropped sets. Second, models are trained on both versions of ICDAR 2017, then tested on the corresponding versions of ICDAR 2013. It is worth to note that testing a model on a test set that is completely different from the training set is challenging for the learning models. Dataset details are presented in the following subsection.

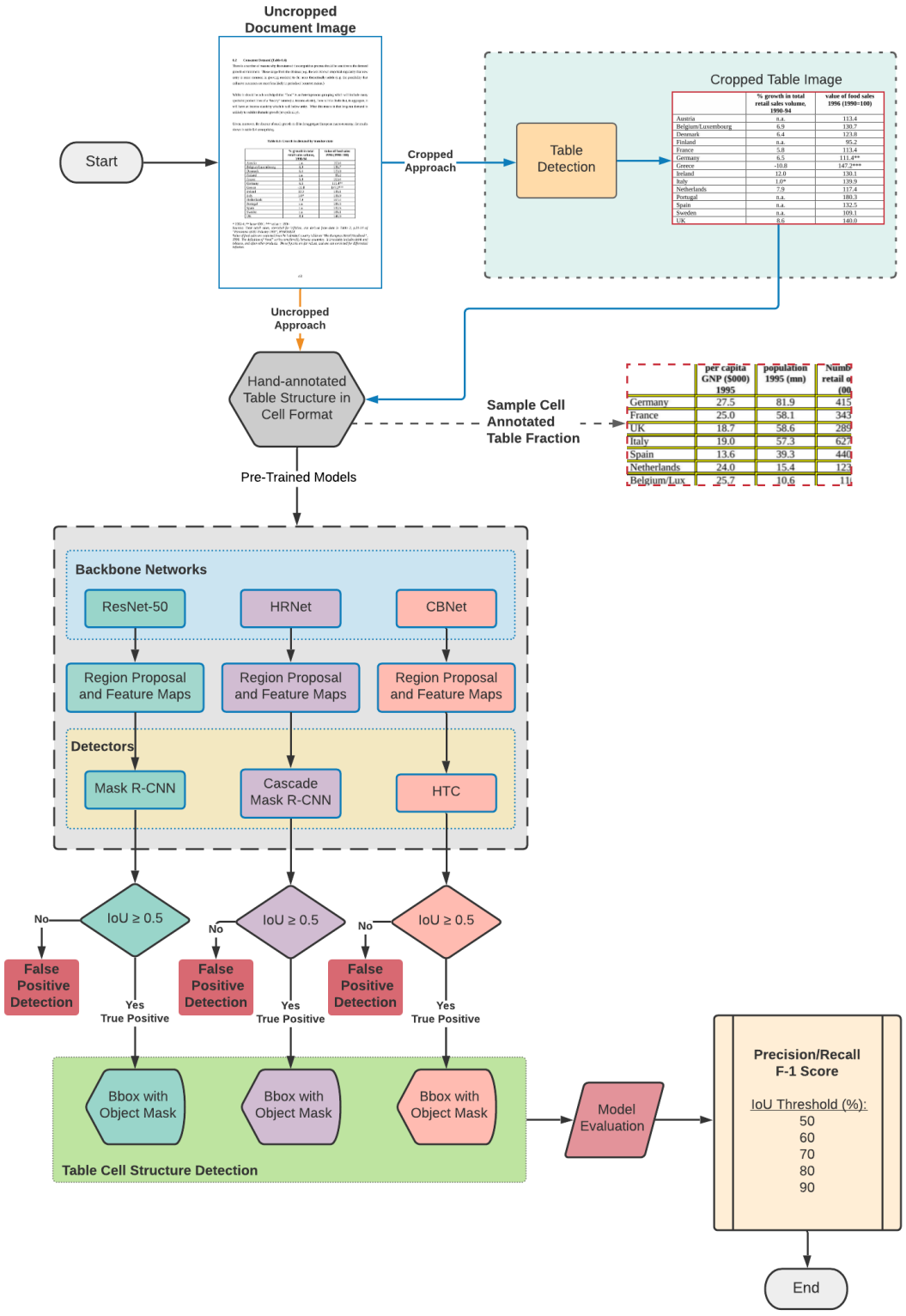


Figure 5.1: Flow diagram of cropped and uncropped approaches

5.2 Datasets

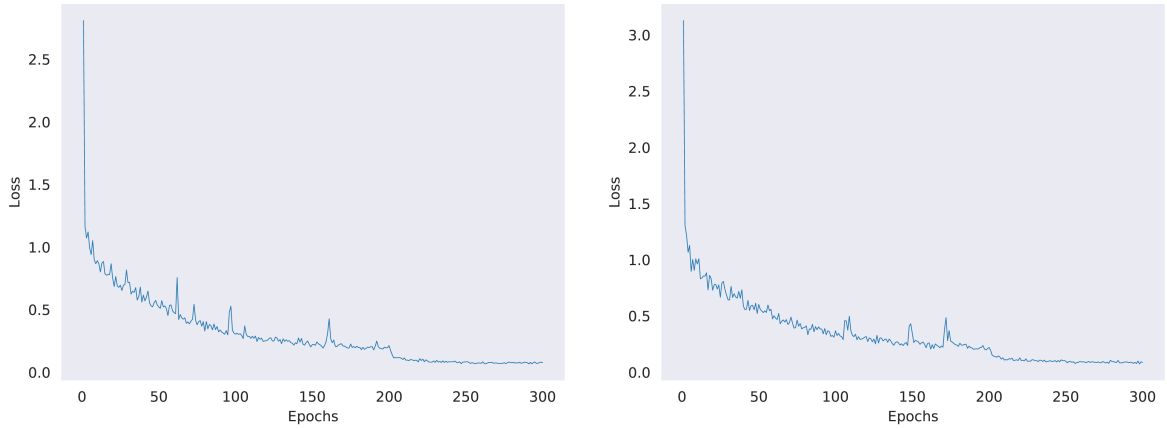
The publicly available ICDAR 2013 [68], and ICDAR 2017 [40] datasets have been widely used in table detection solutions. In this study, we use these datasets with their cropped and uncropped versions. The existing Images that only contain tables are included in both datasets. The competition organizers provide table locations and cell locations of the ICDAR 2013 dataset. For the ICDAR 2017 dataset, however, only table location information is available. Cropped versions of these datasets are created by cropping tables from original documents by using the table locations via a python script. Since structure information is not available for ICDAR 2017 dataset, both ICDAR 2013 and ICDAR 2017 datasets are hand-labelled in cell format for table structure detection as used in [15]. If the table ruling lines exist, they are taken as a reference for cell boundaries. When they do not exist, the aligned bounding boxes are created. The dimensions of cell bounding boxes are defined based on the largest content of each row and column. The other public datasets either do not have both table and table structure annotations or are not in the desired format. Hence we are unable to use them for consistent structure definition. When table structure is defined in rows and columns format, dealing with the spanning rows or columns is challenging [12, 13]. Cell format is chosen due to its simplicity in defining table structure. The number of images in datasets is summarized in Table 5.1. Since some of the images contain multiple tables, cropped sets have more images than uncropped images.

Table 5.1: Numerical details of used datasets

	Cropped		Uncropped	
Dataset	Training	Test	Training	Test
ICDAR 2013	-	156	-	128
ICDAR 2017	1012	-	784	-
ICDAR 2017*	806	206	627	157

5.3 Training Details

This section presents the training details of the copped and uncropped training sets by using the state-of-the-art object detection algorithms that have been proposed for the table structure detection task. There are two different sets of experiments with different datasets. The Mist GPU cluster, which consists of 4 Tesla V 100 GPUs with 32GB VRAM per node, was used in all experiments. All models were implemented by using the mmdetection



(a) Learning Curve for Mask R-CNN on Cropped Set

(b) Learning Curve for Mask R-CNN on Uncropped Set

Figure 5.2: Learning Curves

toolbox [71]. To carry out experiments, Mask R-CNN is chosen as a baseline model. The presented works in [12, 13] has shown that Mask R-CNN performs well on structure detection task. Besides, Mask R-CNN has been widely preferred on instance segmentation due to promising performance and less complex structure. Improving structure detection performance by developing network architecture is not the focus of this study. However, to further investigate the cropped and uncropped sets and justify findings, two more models are used in our experiments. Prasad et al. [14] achieved the highest structure detection results on ICDAR 2019 dataset by using Cascade Mask R-CNN with High-Resolution Network (HRNet) backbone. Jiang et al. [15] created a benchmark study on ICDAR 2013 dataset by using various models with the combination of different backbones. Hybrid Task Cascade (HTC) with double ResNeXt101 outperform all models and shown to be the best candidate for table structure detection. Therefore, Cascade Mask R-CNN HRNet and HTC with dual ResNeXt101 can be chosen for experiments. All models are trained for 300 epochs with both uncropped sets and cropped sets separately. A learning rate of 0.005 was chosen for each model, and training loss and accuracy were watched to assure that the model is fully trained. The learning curves of Mask R-CNN with ResNet-50 for the cropped and uncropped version are presented in Fig. 5.2a and Fig. 5.2b respectively.

5.4 Results

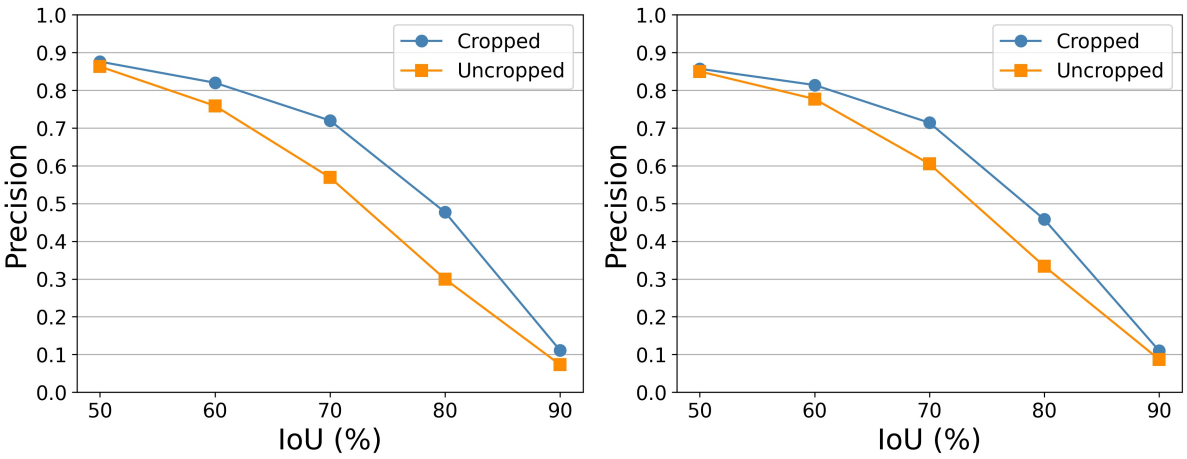
Detections are evaluated based on precision, recall and F1-score along with the IoU concept. IoU concept is explained in the Section 4.4 in detail. The two sets of experiments are discussed in the following sections based on these evaluation metrics.

5.4.1 Evaluation under ICDAR 2017 Dataset

In the first set of experiments, the uncropped ICDAR 2017* dataset is split into training and test sets in 80% and 20%, respectively. Numerical details are given in Table 5.1. The cropped training and test sets are created with the corresponding cropped table images. In other words, 806 table images in the cropped ICDAR 2017 dataset are obtained from 627 images in the uncropped training set by cropping tables. The structure detection results of various networks are presented in Table 5.2, 5.3, and 5.4 for the cropped and uncropped versions. All models achieve 6%-9% higher AP, 5%-8% higher AR and 6%-8% higher AF1 consequently on the cropped set. When the results are examined for different IoU thresholds, an interesting relation is observed. Precision and recall values of models are quite close to each other on the cropped and uncropped sets at 50% IoU. Performance difference between these sets can go up to 17% and 15% of precision and recall values, respectively. These are illustrated with precision-IoU plots in Fig. 5.3 for each model.

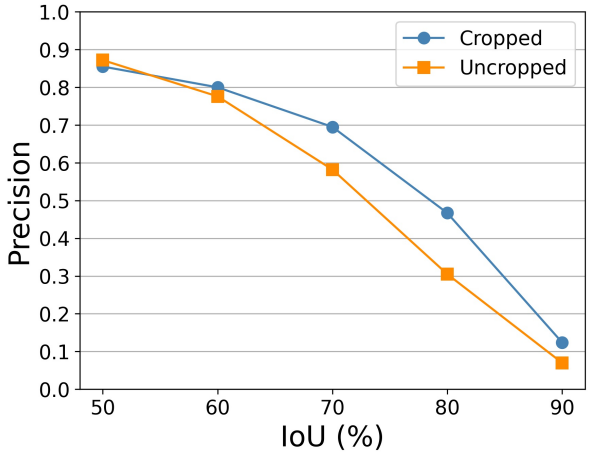
Table 5.2: Structure Detection Results for The Cropped and Uncropped ICDAR 2017 Datasets with Mask R-CNN

			IoU				
Metric	Version	Average	0.50%	0.60%	0.70%	0.80%	0.90%
Precision	Cropped	**0.557**	**0.876**	**0.820**	**0.720**	**0.477**	**0.111**
	Uncropped	0.468	0.863	0.759	0.569	0.300	0.073
Recall	Cropped	**0.665**	**0.929**	**0.890**	**0.818**	**0.639**	**0.279**
	Uncropped	0.588	0.919	0.847	0.713	0.491	0.204
F1-Score	Cropped	**0.606**	**0.902**	**0.854**	**0.766**	**0.546**	**0.159**
	Uncropped	0.521	0.890	0.801	0.633	0.372	0.108



(a) Mask R-CNN

(b) Cascade Mask R-CNN



(c) Hybrid Task Cascade

Figure 5.3: Comparison of precision values of cropped and uncropped ICDAR 2017 datasets with different models

Table 5.3: Structure Detection Results for The Cropped and Uncropped ICDAR 2017 Datasets with Cascade Mask R-CNN

			IoU				
Metric	Version	Average	0.50%	0.60%	0.70%	0.80%	0.90%
Precision	Cropped	**0.546**	**0.857**	**0.814**	**0.714**	**0.458**	**0.110**
	Uncropped	0.487	0.850	0.777	0.605	0.334	0.087
Recall	Cropped	**0.638**	**0.907**	**0.871**	**0.794**	**0.597**	**0.250**
	Uncropped	0.579	0.906	0.854	0.714	0.471	0.181
F1-Score	Cropped	**0.588**	**0.881**	**0.842**	**0.752**	**0.518**	**0.153**
	Uncropped	0.529	0.877	0.814	0.655	0.391	0.118

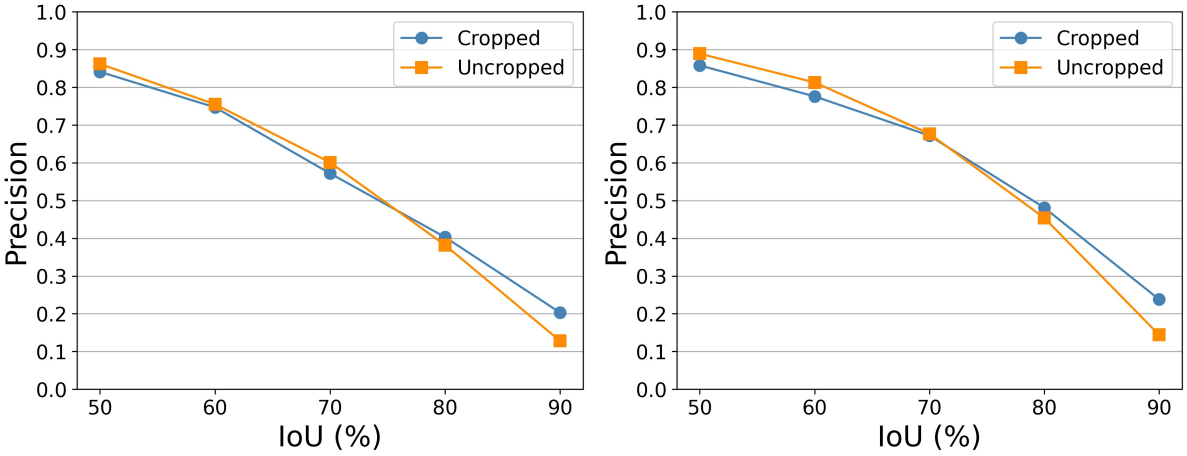
Table 5.4: Structure Detection Results for The Cropped and Uncropped ICDAR 2017 Datasets with Hybrid Task Cascade

			IoU				
Metric	Version	Average	0.50%	0.60%	0.70%	0.80%	0.90%
Precision	Cropped	**0.544**	0.855	**0.800**	**0.695**	**0.467**	**0.124**
	Uncropped	0.476	**0.872**	0.776	0.582	0.305	0.070
Recall	Cropped	**0.646**	0.918	**0.874**	**0.789**	**0.601**	**0.280**
	Uncropped	0.597	**0.932**	0.869	0.727	0.493	0.200
F1-Score	Cropped	**0.591**	0.885	**0.835**	**0.739**	**0.526**	**0.172**
	Uncropped	0.530	**0.901**	0.820	0.646	0.377	0.104

5.4.2 Evaluation under ICDAR 2013 + ICDAR 2017 Datasets

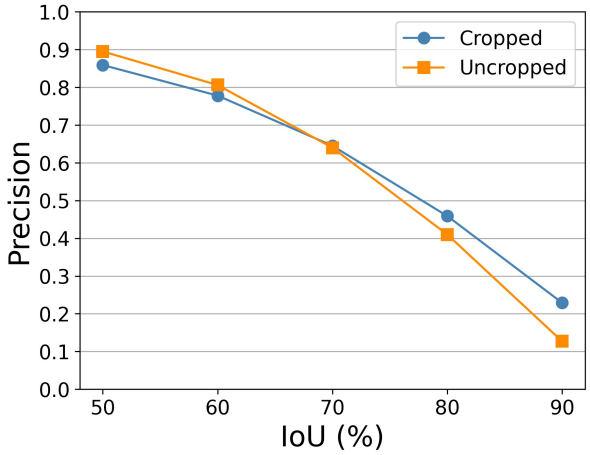
In the second experiment, the cropped and uncropped version of the ICDAR 2017 dataset is used as the training dataset and trained models tested on the corresponding versions of the ICDAR 2013 dataset. In Table 5.5, 5.6, and 5.7, the structure detection results are given with state-of-the-art table structure detection models. When Mask R-CNN is trained with the cropped dataset, it achieves 0.8% higher AP, 1.4% higher AR and 1.1% higher AF1 than the uncropped set. The uncropped dataset provides higher precision and recall values at 50% and lower IoU values up until 80% IoU. As observed in section 5.4.1, the performance of cropped sets increases with the increasing IoUs.

Comparison between cropped and uncropped datasets is further investigated with Cascade Mask R-CNN and HTC. Cascade Mask R-CNN also provides one percent better AP and 0.4% AF1 score under the cropped set, while the average AR is 0.3% lower than the



(a) Mask R-CNN

(b) Cascade Mask R-CNN



(c) Hybrid Task Cascade

Figure 5.4: Comparison of precision values of cropped and uncropped ICDAR 2013 datasets with different models

Table 5.5: Structure Detection Results for The Cropped and Uncropped ICDAR 2013 Datasets with Mask R-CNN

			IoU				
Metric	Version	Average	0.50%	0.60%	0.70%	0.80%	0.90%
Precision	Cropped	**0.509**	0.841	0.747	0.572	**0.403**	**0.203**
	Uncropped	0.501	**0.862**	**0.755**	**0.601**	0.382	0.128
Recall	Cropped	**0.596**	0.889	0.820	0.673	**0.505**	**0.310**
	Uncropped	0.582	**0.894**	**0.820**	**0.696**	0.499	0.232
F1-Score	Cropped	**0.549**	0.864	0.782	0.618	**0.448**	**0.245**
	Uncropped	0.538	**0.878**	**0.786**	**0.645**	0.433	0.165

Table 5.6: Structure Detection Results for The Cropped and Uncropped ICDAR 2013 Datasets with Cascade Mask R-CNN

			IoU				
Metric	Version	Average	0.50%	0.60%	0.70%	0.80%	0.90%
Precision	Cropped	**0.559**	0.858	0.776	0.672	**0.481**	**0.238**
	Uncropped	0.549	**0.889**	**0.813**	**0.677**	0.454	0.144
Recall	Cropped	0.611	0.881	0.819	0.724	0.543	**0.309**
	Uncropped	**0.614**	**0.911**	**0.854**	**0.749**	**0.556**	0.235
F1-Score	Cropped	0.584	0.869	0.797	0.697	**0.510**	**0.269**
	Uncropped	**0.580**	**0.900**	**0.833**	**0.711**	0.500	0.179

Table 5.7: Structure Detection Results for The Cropped and Uncropped ICDAR 2013 Datasets with Hybrid Task Cascade

			IoU				
Metric	Version	Average	0.50%	0.60%	0.70%	0.80%	0.90%
Precision	Cropped	**0.548**	0.859	0.778	**0.645**	**0.459**	**0.229**
	Uncropped	0.529	**0.895**	**0.806**	0.640	0.410	0.127
Recall	Cropped	**0.631**	0.910	0.851	0.736	**0.556**	**0.324**
	Uncropped	0.628	**0.939**	**0.874**	**0.747**	0.554	0.263
F1-Score	Cropped	**0.587**	0.884	0.813	0.688	**0.503**	**0.268**
	Uncropped	0.574	**0.916**	**0.839**	**0.689**	0.471	0.171

uncropped set. Similar to Mask R-CNN, Cascade Mask R-CNN performs well at higher IoUs on the cropped set. The AP under the cropped set is 1.9% improved compared to the uncropped version with the HTC model. The same trend in evaluation metrics between the cropped and uncropped versions for lower and higher IoUs is observed with the HTC model as well. This trend can be seen in Fig. 5.4 for each model. The performance difference between the cropped and uncropped sets is relatively less evident on the whole ICDAR 2013 than ICDAR 2017 test sets. Although it requires further investigation, it is possible that this is due to the fact that the training and test sets come from different distributions.

Overall, the structure detection performance of all models is improved on the cropped table images for AP, AR and AF1 metrics. Models either provided better cell structure detection based on the evaluation metrics under lower IoU values on the uncropped table images or performed similarly with the cropped sets. However, with the increasing IoU thresholds, models achieve significantly higher AP, AR and AF1 values on cropped sets. In other words, cell structures are detected more accurately on cropped table images. Cell structure detection samples are presented in Fig. 5.5 for uncropped table images and corresponding cropped table images.

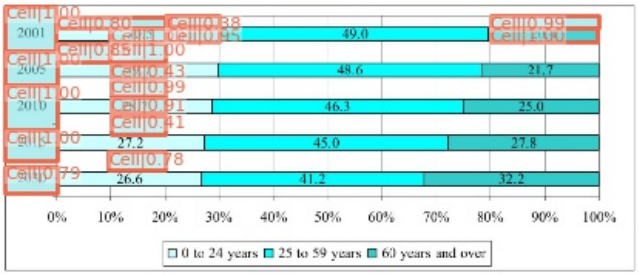
To complement numerical results in the study, figure 5.5 provides sample inference results on ICDAR2013 dataset. Figure 5.5a and 5.5c show that lines in chart figures within the documents cause false positive detections. On croppped table images, these false detections can be eliminated as seen in Figure 5.5b and 5.5d.

Furthermore, vertical or horizontal alignments of the textual region might be perceived as tabular information. By working on cropped sets, these deceiving factors can be reduced to minimum. A sample image that shows false positive cell object detections in a textual area can be seen in Figure 5.5e and the corresponding cropped table structure detections in 5.5f.

5.5 Summary

Despite lacking analysis on the interplay between them, table detection and table structure recognition are have been considered two consecutive tasks. In this chapter, we have investigated the table cell structure detection performance on the cropped and uncropped versions of the ICDAR 2013 and ICDAR 2017 datasets. Comparison of these two versions illustrates the impact of having a table detection model. By proving that the cropped version provides remarkably better performance, it has become a guide for researchers in

Figure 1: Population by age group in 2001, 2005 and forecasts for 2010, 2015 and 2030 (in %)



Source: Statistics Finland, 2006.

1.3. The economy and the labour market

The economy and welfare have grown steadily in Finland since independence until the 1990s, except during the depression in the 1930s and the Second World War. In the 1950s, trade with the Soviet Union had a significant impact on the development of export industries. First the war indemnities to Soviet Union and then the bilateral trade relations with it meant a rapid increase in industrial activity in Finland. In the 1980s growth was stable but, at the beginning of the 1990s, the Finnish national economy was hit by the worst depression since the war. The growth of GDP in recent years has been faster than in the EU in general (see Table 1).

Table 1: Real GDP growth rate in Finland, EU-15 and EU-25 for 1996, 2000, 2005 and 2006 (percentage change on previous year)

	Finland	EU-15	EU-25
1996	3.7	1.6	1.7
2000	5.0	3.9	3.9
2005	1.5	1.5	1.6
2006 (*)	3.5	2.0	2.1

GDP: Gross domestic product.
(*) Forecast.
Source: Eurostat. European system of accounts (ESA 1995), 2005.

Finland has the industrial structure of a modern knowledge-based society. The proportion of agriculture and manufacturing has declined and, in the last two decades, electronics has become the success story of Finnish exports. Its growth in the 1990s is mainly based on mobile phones and other telecommunication equipment. Three major export sectors today are

(a)

	Finland	EU-15	EU-25
1996	3.7	1.6	1.7
2000	5.0	3.9	3.9
2005	1.5	1.5	1.6
2006 (*)	3.5	2.0	2.1

(b)

CESR

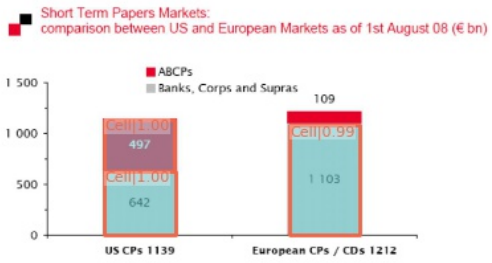
135. Specific events and factors were of particular importance in the decline of ABCPs. Firstly, some conduits had large ABS holdings that experienced huge declines. When investors stopped rolling over ABCPs, these conduits had to rely on guarantees provided by banks which were too large for the banks providing them. While these banks received support to meet their obligations, investor confidence was nonetheless damaged. Secondly, structures in other ABCP markets around the world unsettled investors, including different guarantee agreements and single-seller extendible mortgage conduits. Thirdly, general concerns about the banking sector have caused investors to buy less bank related product.

Table 3 - European ABCP issuance

	Q1	Q2	Q3	Q4	Total
2004	34.7	36.2	44.3	51.5	166.7
2005	58.1	63.4	61.6	55.2	238.4
2006	74.7	84.1	96.3	111.8	367.1
2007	148.8	142.3	156.7	186.1	633.9
2008	120.9	106			226.8

Source: Moody's, Dealogic, ESF

Chart 5



Source: Société Générale Corporate & Investment Banking (market overview, 19 September, 2008)

Credit Derivatives Markets

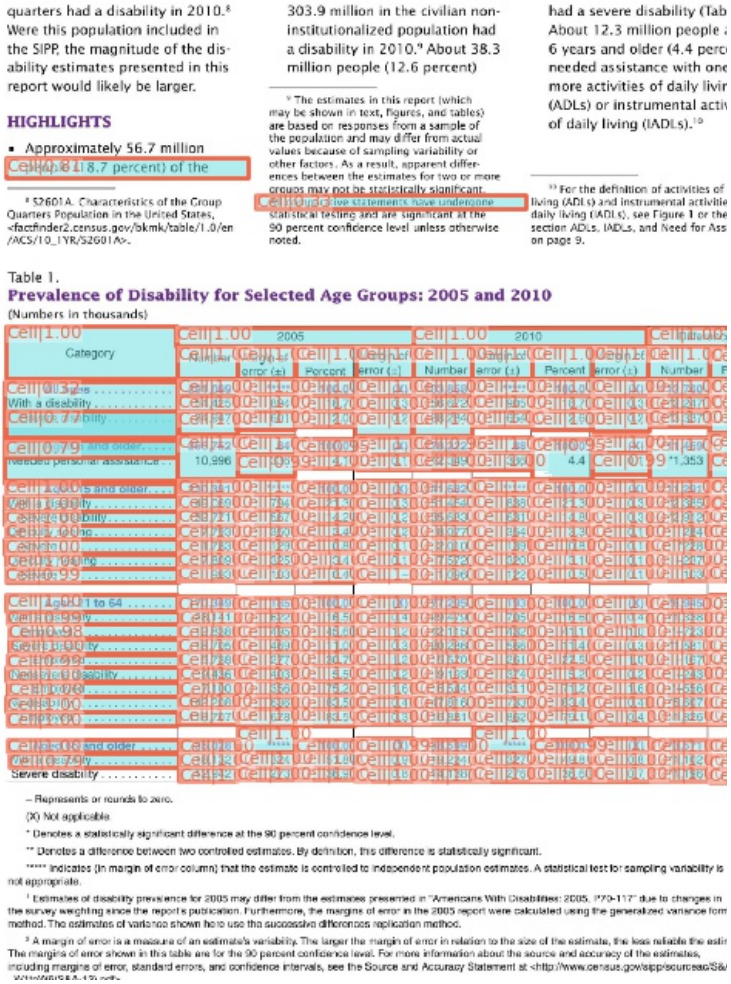
136. The credit derivatives markets comprise a number of instruments. Credit default swaps represent, by far, the single most significant credit derivative instrument in terms of volume. Other credit derivative instruments are not covered in this consultation paper[43].

[43] Examples of credit derivatives not included in the scope of this consultation paper are total return swaps and credit linked notes.

(c)

	Q1	Q2	Q3	Q4	Total
2004	34.7	36.2	44.5	51.3	166.7
2005	58.1	63.4	61.6	55.2	238.4
2006	74.7	84.1	96.5	111.8	367.1
2007	148.8	142.3	156.7	186.1	633.9
2008	120.9	106			226.8

(d)



quarters had a disability in 2010.[8] Were this population included in the SIPP, the magnitude of the disability estimates presented in this report would likely be larger.

303.9 million in the civilian noninstitutionalized population had a disability in 2010.* About 38.3 million people (12.6 percent)

had a severe disability (Tab
About 12.3 million people
6 years and older (4.4 perc
needed assistance with one
more activities of daily livin
(ADLs) or instrumental activ
of daily living (IADLs).[10]

HIGHLIGHTS

- Approximately 56.7 million

[8] S2601A. Characteristics of the Group Quarters Population in the United States, <factfinder2.census.gov/bkmk/table/1.0/en/ACS/10_1YR/S2601A>.

* The estimates in this report (which may be shown in text, figures, and tables) are based on responses from a sample of the population and may differ from actual values because of sampling variability or other factors. As a result, apparent differences between the estimates for two or more groups may not be statistically significant. All the statements about differences given in this report have undergone statistical testing and are significant at the 90 percent confidence level unless otherwise noted.

[10] For the definition of activities of living (ADLs) and instrumental activities daily living (IADLs), see Figure 1 or the section ADLs, IADLs, and Need for Ass on page 9.

Table 1.
Prevalence of Disability for Selected Age Groups: 2005 and 2010
(Numbers in thousands)

Category	2005			2010			Difference	
	Number	Margin of error (±)	Percent	Number	Margin of error (±)	Percent	Number	Percent
Severe disability								
Aged 6 and older	10,996	336	4.1	12,349	388	4.4	*1,353	*0.3
Aged 15 and older								
With a disability								
Severe disability								
Difficulty seeing								
Severe								
Difficulty hearing								
Severe								
Aged 21 to 64								
With a disability								
Employed								
Severe disability								
Employed								
Nonsevere disability								
Employed								
No disability								
Employed								
Aged 65 and older								
With a disability								
Severe disability								

– Represents or rounds to zero.
(X) Not applicable.
* Denotes a statistically significant difference at the 90 percent confidence level.
** Denotes a difference between two controlled estimates. By definition, this difference is statistically significant.
***** Indicates (in margin of error column) that the estimate is controlled to independent population estimates. A statistical test for sampling variability is not appropriate.

[1] Estimates of disability prevalence for 2005 may differ from the estimates presented in "Americans With Disabilities: 2005, P70-117" due to changes in the survey weighting since the report's publication. Furthermore, the margins of error in the 2005 report were calculated using the generalized variance formula method. The estimates of variance shown here use the successive differences replication method.

[3] A margin of error is a measure of an estimate's variability. The larger the margin of error in relation to the size of the estimate, the less reliable the estimate. The margins of error shown in this table are for the 90 percent confidence level. For more information about the source and accuracy of the estimates, including margins of error, standard errors, and confidence intervals, see the Source and Accuracy Statement at <http://www.census.gov/sipp/sourceacc/S&

(e)

Category	2005			2010			Difference	
	Number	Margin of error (±)	Percent	Number	Margin of error (±)	Percent	Number	Percent
All ages	291,099	*****	100.0	(X)	303,858	*****	100.0	(X)	**12,760	(X)
With a disability	54,425	894	18.7	0.3	56,672	905	18.7	0.3	**2,247	–
Severe disability	34,947	601	12.0	0.2	39,284	654	12.6	0.2	**4,337	0.6
Aged 6 and older	266,752	84	100.0	(X)	278,222	88	100.0	(X)	**11,469	(X)
Needed personal assistance	10,996	336	4.1	0.1	12,349	388	4.4	0.1	*1,353	*0.3
Aged 15 and older	230,391	*****	100.0	(X)	241,682	*****	100.0	(X)	**11,291	(X)
With a disability	49,069	794	21.3	0.3	51,454	838	21.3	0.3	**2,385	–
Severe disability	32,771	567	14.2	0.2	35,683	631	14.8	0.3	**2,912	0.5
Difficulty seeing	7,793	350	3.4	0.2	8,077	354	3.3	0.1	284	–
Severe	1,783	129	0.8	0.1	2,010	139	0.8	0.1	228	0.1
Difficulty hearing	7,809	325	3.4	0.1	7,572	320	3.1	0.1	**237	**0.3
Severe	993	103	0.4	–	1,096	122	0.5	0.1	103	–
Aged 21 to 64	170,349	185	100.0	(X)	177,295	193	100.0	(X)	**6,945	(X)
With a disability	28,141	622	16.5	0.4	29,479	705	16.6	0.4	**1,338	0.1
Employed	12,838	495	45.6	1.2	12,115	432	41.1	1.0	**723	**4.5
Severe disability	15,705	469	11.0	0.3	20,286	566	11.4	0.3	**4,581	**0.5
Employed	5,738	277	30.7	1.2	5,570	261	27.5	1.0	**167	**3.2
Nonsevere disability	9,436	403	5.5	0.2	9,193	374	5.2	0.2	**243	**0.4
Employed	7,100	356	75.2	1.6	6,544	311	71.2	1.6	**556	**4.1
No disability	142,208	636	83.5	0.4	147,816	733	83.4	0.4	**5,607	**0.1
Employed	118,707	678	83.5	0.3	110,881	862	79.1	0.4	**7,826	**4.4
Aged 65 and older	35,028	*****	100.0	(X)	38,599	*****	100.0	(X)	**3,571	(X)
With a disability	18,132	324	51.8	0.9	19,234	327	49.8	0.8	**1,102	**1.9
Severe disability	12,942	273	36.9	0.8	14,138	276	36.6	0.7	**1,196	**0.3

(f)

Figure 5.5: Sample structure detection results on cropped and uncropped ICDAR 2013 datasets.

future table structure detection studies. Experiments are initially done by using Mask R-CNN. Cascade Mask R-CNN and Hybrid Task Cascade are used for further analysis. Models achieved higher AP and AR on the cropped datasets. Besides, the following impact of the IoU thresholds on model performance has been been discovered: Models can provide better detections at lower IoU values of 50%-70% on either set, while they perform noticeably better under higher IoU values on the cropped set. The performance gap can be as wide as 15% and 17% in terms of the precision and recall values. It can be concluded that none-table objects in document images such as textual areas, figures and plots degrade the model performance for higher IoU thresholds. Hence a robust table detection model improves the performance of a table structure detection model. Finally, false-positive detections due to lines existing in figures or alignments in textual areas can be eliminated by the cropping process.

Chapter 6

Conclusion

The documents are generated in very high volumes and varying formats through various channels. It is not very feasible to process these documents with human power in an efficient and effective manner. However, retrieving useful data from documents is highly desired by a range of organizations from industry to government. Table analysis is an essential part of document processing since valuable information is often presented in tables within the documents. Many studies have been conducted to obtain tabular data in an automated manner. Traditional approaches that rely on heuristics and pre-defined rules lack the ability to generalize because the structure and layouts of tables vary greatly. Deep learning approaches show great promise and can be applied in a variety of formats and any context. There are two main stages of obtaining information from a table, first is detecting and localizing the table; the second is segmenting the table into its fundamental components.

The problem of table detection is an extensively studied problem and a popular research subject in document analysis. The proposed deep learning methods[72, 11] provide excellent results and are very close to the saturation point in terms of performance. The table structure detection stage of tabular information extraction, on the other hand, is significantly open to development and has far too many issues to address. In this thesis, we assume that the table detection step is realized flawlessly and focus on the table structure detection problem. Deep learning-based solutions that adapt R-CNNs proved to be the most effective and reliable approach. They have more generalization capability and are format-independent since PDF documents can easily be converted to images without requiring metadata. First, we aim to make the table definition in the most optimal way that reduces the complexity of the problem and makes it less error-prone. To achieve this, we propose a content-focused cell detection method where the table elements are defined in cell format and the boundary of bounding boxes are limited to the text, character or, in

some cases, even figure content. Thus, the performance decrease caused by the ambiguities due to the large white spaces and lack of ruling lines is avoided. This approach provides a 93.9% of average F1-score at the 50% IoU threshold and a 73.8% average F1-score without bells and whistles. A base model that is Mask R-CNN and ResNet101 as a backbone is used to analyze optimal table structure definition since we are not focusing on developing network structure in this task. Besides, achieving 100% overlapping between the ground-truth and predicted object is challenging even in general object detection tasks. Following the optimal table structure definition, training schemes of structure detection models are investigated. In the uncropped training scheme, models are trained with the regular document images where non-table objects such as figures, graphs and textual regions appear. In the cropped training scheme, models are trained with cropped table images. Mask R-CNN, Cascade Mask R-CNN and Hybrid Task Cascade are implemented for the analysis of both approaches. Models perform interchangeably better with each other at low IoU values of 50%-70%, while the cropped training scheme yields remarkably better results. The cropped approach can achieve higher precision and recall values up to 17% and 15%, respectively. Overall, this thesis proposes a content-focused cell detection method with cropped training scheme.

6.1 Future Work

Our proposed method can achieve around 95% precision and recall values at 50% IoU and slightly decreasing performance with higher IoU thresholds that is natural to occur. However, there is still room for improvement in structure detection performance at higher IoU values where performance degrades significantly. A Large-scale dataset annotated with table structure information is an absolute must since deep learning models are data-driven. Recently released large-scale datasets are not applicable to all approaches for a variety of reasons, including differences in table structure definitions, domain-specific datasets, and synthetic datasets. A critical issue is that table structures might vary significantly even within the same domain. Although the content-focused cell detection method handles many variations, it introduces new challenges; there may be a large number of objects, and they may be extraordinarily small compared to row/column. The challenges and issues that we addressed in this thesis can still be considered for further improvements along with other possible solutions.

Large-scale dataset annotated with table structure information

The ICDAR2013 and ICDAR2017 datasets have been annotated in various table structure formats, but these need to be extended to larger datasets annotated with the content-focused cell approach. Publicly available datasets such as ICDAR2019 [73], PubTabNet [51] and SciTSR [48] can be annotated, or existing annotations can be converted accordingly for a more generalized table structure detection.

Small object detection in densely packed doucument images

Target objects in the content-focused cell approach are extremely small in size, where sometimes it covers only a single character as table content. Also, they appear densely and in very close proximity on a single document image. These challenges can be addressed with network-level improvements. Therefore adopting a detection network that addresses the problem of small object detection and densely packed object detection is a critical focus of research direction for better table structure detection.

Structure detection on clustered or classified tables

Table layouts do not have a universal standard, and they come in a variety of styles. This variation prevents the better generalization of deep learning models on table structure detection. A clustering approach can minimize the heterogeneity of table objects prior to the structure detection step. Several structure detection models can be used after determining an optimal number of clusters, each specialized on a particular cluster. Another alternative solution can be classifying tables with pre-determined table classes in a supervised manner. The presence of ruling lines or the homogeneity of table structure might be used as classification criteria. From the point of view of efficiency, unsupervised methods are more attractive because there is no need for data processing. However, their performance may be suggestive since distinguishing tabular objects is not a trivial task.

Vision transformers in structure detection

CNNs are the driving methods in object detection and instance segmentation. Astounding achievements of transformer models on natural language tasks have sparked their applications to computer vision problems. Beyond their promising performance, the design of

a self-attention mechanism in multiple heads allows processing multiple modalities. Structure detection performance can be significantly improved by learning visual representations and text data simultaneously.

References

[1] O. Ercan and G. Samet, "Literature review of industry 4.0 and related technologies," *Journal of Intelligent Manufacturing*, vol. 31, pp. 127–182, 01 2020. Copyright - Journal of Intelligent Manufacturing is a copyright of Springer, (2018). All Rights Reserved; Last updated - 2020-11-17.

[2] M. Attaran, "Digital technology enablers and their implications for supply chain management," *Supply Chain Forum: An International Journal*, vol. 21, no. 3, pp. 158–172, 2020.

[3] E. Oro and M. Ruffolo, "Trex: An approach for recognizing and extracting tables from pdf documents," in *Intl. Conf. on Document Analysis and Recognition*, pp. 906–910, IEEE, 2009.

[4] J. Fang, L. Gao, K. Bai, R. Qiu, X. Tao, and Z. Tang, "A table detection method for multipage pdf documents via visual seperators and tabular structures," in *2011 International Conference on Document Analysis and Recognition*, pp. 779–783, 2011.

[5] J. Fang, L. Gao, K. Bai, R. Qiu, X. Tao, and Z. Tang, "A table detection method for multipage pdf documents via visual seperators and tabular structures," in *2011 International Conference on Document Analysis and Recognition*, pp. 779–783, IEEE, 2011.

[6] D. N. Tran, T. A. Tran, A. Oh, S. H. Kim, and I. S. Na, "Table detection from document image using vertical arrangement of text blocks," *International Journal of Contents*, vol. 11, no. 4, pp. 77–85, 2015.

[7] M. Traquair, E. Kara, B. Kantarci, and S. Khan, "Deep learning for the detection of tabular information from electronic component datasheets," in *IEEE Symposium on Computers and Communications (ISCC)*, (Barcelona, Spain), June 2019.

[8] F. Cesarini, S. Marinai, L. Sarti, and G. Soda, "Trainable table location in document images," in *Object recognition supported by user interaction for service robots*, vol. 3, pp. 236–240, IEEE, 2002.

[9] L. Hao, L. Gao, X. Yi, and Z. Tang, "A table detection method for pdf documents based on convolutional neural networks," in *2016 12th IAPR Workshop on Document Analysis Systems (DAS)*, pp. 287–292, IEEE, 2016.

[10] A. Gilani, S. R. Qasim, I. Malik, and F. Shafait, "Table detection using deep learning," in *2017 14th IAPR international conference on document analysis and recognition (ICDAR)*, vol. 1, pp. 771–776, IEEE, 2017.

[11] J. Jiang, M. Simsek, B. Kantarci, and S. Khan, "High precision deep learning based tabular detection," in *IEEE Symposium on Computers and Communications (ISCC)*, (Rennes, France), 2020.

[12] E. Kara, M. Traquair, B. Kantarci, and S. Khan, "Deep learning for recognizing the anatomy of tables on datasheets," in *2019 IEEE Symposium on Computers and Communications (ISCC)*, pp. 1–6, 2019.

[13] E. Kara, M. Traquair, M. Simsek, B. Kantarci, and S. Khan, "Holistic design for deep learning-based discovery of tabular structures in datasheet images," *Engineering Applications of Artificial Intelligence*, vol. 90, p. 103551, 2020.

[14] D. Prasad, A. Gadpal, K. Kapadni, M. Visave, and K. Sultanpure, "Cascadetabnet: An approach for end to end table detection and structure recognition from image-based documents," in *Proceedings of the IEEE/CVF Conference on CVPR Workshops*, pp. 572–573, 2020.

[15] J. Jiang, M. Simsek, B. Kantarci, and S. Khan, "Tabcellnet: Deep learning-based tabular cell structure detection," *Neurocomputing*, 2021.

[16] J. Singer-Vine, "pdfplumber." https://github.com/jsvine/pdfplumber, 2015.

[17] R. Girshick, "Fast r-cnn," in *IEEE Intl. Conf. on Computer Vision*, pp. 1440–1448, 2015.

[18] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.

[19] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *IEEE Intl. Conf. on Computer Vision*, pp. 2980–2988, 2017.

[20] Z. Cai and N. Vasconcelos, "Cascade r-cnn: High quality object detection and instance segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2019.

[21] K. Chen, J. Pang, J. Wang, Y. Xiong, X. Li, S. Sun, W. Feng, Z. Liu, J. Shi, W. Ouyang, C. C. Loy, and D. Lin, "Hybrid task cascade for instance segmentation," in *IEEE/CVF CVPR*, pp. 4969–4978, 2019.

[22] K. A. Hashmi, M. Liwicki, D. Stricker, M. A. Afzal, M. A. Afzal, and M. Z. Afzal, "Current status and performance analysis of table recognition in document images with deep neural networks," *IEEE Access*, vol. 9, pp. 87663–87685, 2021.

[23] M. Hearst, S. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intelligent Systems and their Applications*, vol. 13, no. 4, pp. 18–28, 1998.

[24] J. R. Quinlan, "Induction of decision trees," *Machine learning*, vol. 1, no. 1, pp. 81–106, 1986.

[25] K. Zuyev, "Table image segmentation," in *Proceedings of the Fourth International Conference on Document Analysis and Recognition*, vol. 2, pp. 705–708 vol.2, 1997.

[26] P. Pyreddy and W. B. Croft, "Tintin: A system for retrieval in text tables," in *Proceedings of the Second ACM International Conference on Digital Libraries*, DL '97, (New York, NY, USA), p. 193–200, Association for Computing Machinery, 1997.

[27] T. Kieninger and A. Dengel, "The t-recs table recognition and analysis system," in *Document Analysis Systems: Theory and Practice* (S.-W. Lee and Y. Nakano, eds.), (Berlin, Heidelberg), pp. 255–270, Springer Berlin Heidelberg, 1999.

[28] A. Pivk, P. Cimiano, Y. Sure, M. Gams, V. Rajkovič, and R. Studer, "Transforming arbitrary tables into logical form with tartar," *Data & Knowledge Engineering*, vol. 60, no. 3, pp. 567–595, 2007.

[29] Y. Wang, I. T. Phillips, and R. M. Haralick, "Table structure understanding and its performance evaluation," *Pattern recognition*, vol. 37, no. 7, pp. 1479–1497, 2004.

[30] J. Ha, R. Haralick, and I. Phillips, "Recursive x-y cut using bounding boxes of connected components," in *Proceedings of 3rd International Conference on Document Analysis and Recognition*, vol. 2, pp. 952–955 vol.2, 1995.

[31] E. Oro and M. Ruffolo, "Pdf-trex: An approach for recognizing and extracting tables from pdf documents," in *2009 10th International Conference on Document Analysis and Recognition*, pp. 906–910, 2009.

[32] T. Hassan and R. Baumgartner, "Table recognition and understanding from pdf files," in *Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)*, vol. 2, pp. 1143–1147, 2007.

[33] S. Khusro, A. Latif, and I. Ullah, "On methods and tools of table detection, extraction and annotation in pdf documents," *Journal of Information Science*, vol. 41, no. 1, pp. 41–57, 2015.

[34] 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*, pp. 1185–1189, 2013.

[35] H. T. Ng, C. Y. Lim, and J. L. T. Koo, "Learning to recognize tables in free text," in *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, pp. 443–450, 1999.

[36] S. F. Rashid, A. Akmal, M. Adnan, A. A. Aslam, and A. Dengel, "Table recognition in heterogeneous documents using machine learning," in *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, vol. 01, pp. 777–782, 2017.

[37] J. Bhatt, K. A. Hashmi, M. Z. Afzal, and D. Stricker, "A survey of graphical page object detection with deep neural networks," *Applied Sciences*, vol. 11, no. 12, 2021.

[38] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, pp. 1097–1105, 2012.

[39] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, *et al.*, "Imagenet large scale visual recognition challenge," *International journal of computer vision*, vol. 115, no. 3, pp. 211–252, 2015.

[40] L. Gao, X. Yi, Z. Jiang, L. Hao, and Z. Tang, "Icdar2017 competition on page object detection," in *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, vol. 01, pp. 1417–1422, 2017.

[41] S. Schreiber, S. Agne, I. Wolf, A. Dengel, and S. Ahmed, "Deepdesrt: Deep learning for detection and structure recognition of tables in document images," in *Intl. Conf. on Document Analysis and Recognition*, vol. 01, pp. 1162–1167, Nov 2017.

[42] S. A. Siddiqui, I. A. Fateh, S. T. R. Rizvi, A. Dengel, and S. Ahmed, "Deeptabstr: Deep learning based table structure recognition," in *Intl. Conf. on Document Analysis and Recognition*, pp. 1403–1409, 2019.

[43] S. S. Paliwal, V. D, R. Rahul, M. Sharma, and L. Vig, "Tablenet: Deep learning model for end-to-end table detection and tabular data extraction from scanned document images," in *Intl. Conf. on Document Analysis and Recognition*, pp. 128–133, 2019.

[44] L. Qiao, Z. Li, Z. Cheng, P. Zhang, S. Pu, Y. Niu, W. Ren, W. Tan, and F. Wu, "Lgpma: Complicated table structure recognition with local and global pyramid mask alignment," 2021.

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

[46] D. Wei, H. Lu, Y. Zhou, and K. Chen, "Image-based table cell detection: a novel table structure decomposition method with new dataset," in *2020 25th International Conference on Pattern Recognition (ICPR)*, pp. 1–7, 2021.

[47] Y. Zou and J. Ma, "A deep semantic segmentation model for image-based table structure recognition," in *2020 15th IEEE International Conference on Signal Processing (ICSP)*, vol. 1, pp. 274–280, 2020.

[48] S. Raja, A. Mondal, and C. V. Jawahar, "Table structure recognition using top-down and bottom-up cues," in *Computer Vision – ECCV 2020* (A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, eds.), (Cham), pp. 70–86, Springer International Publishing, 2020.

[49] X. Zheng, D. Burdick, L. Popa, X. Zhong, and N. X. R. Wang, "Global table extractor (gte): A framework for joint table identification and cell structure recognition using visual context," in *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 697–706, 2021.

[50] K. A. Hashmi, D. Stricker, M. Liwicki, M. N. Afzal, and M. Z. Afzal, "Guided table structure recognition through anchor optimization," *IEEE Access*, vol. 9, pp. 113521–113534, 2021.

[51] X. Zhong, E. ShafieiBavani, and A. Jimeno Yepes, "Image-based table recognition: Data, model, and evaluation," in *Computer Vision – ECCV 2020* (A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, eds.), (Cham), pp. 564–580, Springer International Publishing, 2020.

[52] B. Smock, R. Pesala, and R. Abraham, "Pubtables-1m: Towards a universal dataset and metrics for training and evaluating table extraction models," 2021.

[53] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580–587, 2014.

[54] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results." http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html.

[55] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255, Ieee, 2009.

[56] A. M. Hafiz and G. M. Bhat, "A survey on instance segmentation: state of the art," *International journal of multimedia information retrieval*, vol. 9, no. 3, pp. 171–189, 2020.

[57] S. Minaee, Y. Y. Boykov, F. Porikli, A. J. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2021.

[58] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.

[59] Z. Cai and N. Vasconcelos, "Cascade r-cnn: Delving into high quality object detection," in *IEEE/CVF Conference on CVPR*, pp. 6154–6162, 2018.

[60] T. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *IEEE Conf. on CVPR*, pp. 936–944, 2017.

[61] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *3rd International Conference on Learning Representations, ICLR 2015,San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings* (Y. Bengio and Y. LeCun, eds.), 2015.

[62] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[63] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[64] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 5987–5995, 2017.

[65] J. Wang, K. Sun, T. Cheng, B. Jiang, C. Deng, Y. Zhao, D. Liu, Y. Mu, M. Tan, X. Wang, W. Liu, and B. Xiao, "Deep high-resolution representation learning for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2020.

[66] Y. Liu, Y. Wang, S. Wang, T. Liang, Q. Zhao, Z. Tang, and H. Ling, "Cbnet: A novel composite backbone network architecture for object detection," *AAAI Conf. on Artificial Intelligence*, vol. 34, pp. 11653–11660, Apr. 2020.

[67] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *European conference on computer vision*, pp. 740–755, Springer, 2014.

[68] D. Karatzas, F. Shafait, S. Uchida, M. Iwamura, L. G. i. Bigorda, S. R. Mestre, J. Mas, D. F. Mota, J. A. Almazàn, and L. P. de las Heras, "ICDAR robust reading competition," in *Intl. Conf. on Document Analysis and Recognition*, pp. 1484–1493, 2013.

[69] Y. Wu, A. Kirillov, F. Massa, W.-Y. Lo, and R. Girshick, "Detectron2." https://github.com/facebookresearch/detectron2, 2019.

[70] P. Goyal, P. Dollár, R. Girshick, P. Noordhuis, L. Wesolowski, A. Kyrola, A. Tulloch, Y. Jia, and K. He, "Accurate, large minibatch sgd: Training imagenet in 1 hour," *arXiv preprint arXiv:1706.02677*, 2017.

[71] K. Chen, J. Wang, J. Pang, Y. Cao, Y. Xiong, X. Li, S. Sun, W. Feng, Z. Liu, J. Xu, Z. Zhang, D. Cheng, C. Zhu, T. Cheng, Q. Zhao, B. Li, X. Lu, R. Zhu, Y. Wu, J. Dai, J. Wang, J. Shi, W. Ouyang, C. C. Loy, and D. Lin, "MMDetection: Open mmlab detection toolbox and benchmark," *arXiv preprint arXiv:1906.07155*, 2019.

[72] J. Fernandes, M. Simsek, B. Kantarci, and S. Khan, "Tabledet: An end-to-end deep learning approach for table detection and table image classification in data sheet images," *Neurocomputing*, vol. 468, pp. 317–334, 2022.

[73] L. Gao, Y. Huang, H. Déjean, J.-L. Meunier, Q. Yan, Y. Fang, F. Kleber, and E. Lang, "Icdar 2019 competition on table detection and recognition (ctdar)," in *2019 International Conference on Document Analysis and Recognition (ICDAR)*, pp. 1510–1515, 2019.