Weakly supervised bounding box extraction for unlabeled data in table detection

Arash Samari[1,⋆], Andrew Piper[2], Alison Hedley[3], and Mohamed Cheriet[4]

[1] Ecole De Technologie Supérieure, Montreal, Canada
arash.samari.1.@etsmtl.net
[2] McGill University, Montreal, Canada
andrew.piper@mcgill.ca
[3] McGill University, Montreal, Canada
afhedley@gmail.com
[4] Ecole De Technologie Supérieure, Montreal, Canada
mohamed.cheriet@etsmtl.ca

Abstract. The organization and presentation of data in tabular format became an essential strategy of scientific communication and remains fundamental to the transmission of knowledge today. The use of automated detection to identify typographical elements such as tables and diagrams in digitized historical print offers a promising approach for future research. Most of the table detection tasks are using existing off-the-shelf methods for their detection algorithm. However, datasets that are used for evaluation are not challenging enough due to the lack of quantity and diversity. To have a better comparison between proposed methods we introduce the NAS dataset in this paper for historical digitized images. Tables in historic scientific documents vary widely in their characteristics. They also appear alongside visually similar items, such as maps, diagrams, and illustrations. We address these challenges with a multi-phase procedure, outlined in this article, evaluated using two datasets, ECCO [5] and NAS [6] In our approach, we utilized the Gabor filter [1] to prepare our dataset for algorithmic detection with Faster-RCNN [2]. This method detects tables against all categories of visual information. Due to the limitation in labeled data, particularly for object detection, we developed a new method, namely, weakly supervision bounding box extraction, to extract bounding boxes automatically for our training set in an innovative way. Then a pseudo-labeling technique is used to create a more general model, via a three-step process of bounding box extraction and labeling.

Keywords: Weakly supervision bounding box extraction · Faster-RCNN · pseudo-labeling · Gabor filter · Distance transform.

⋆ Corresponding author

[5] https://www.gale.com/primary-sources/eighteenth-century-collections-online.

[6] https://beta.synchromedia.ca/vok-visibility-of-knowledge.

1 Introduction

In the context of object detection in digitized documents, similar to many other machine learning applications, learning that generalizes well to unseen data is a significant dilemma that we need to address. Moreover, to have a fair comparison between other methods, we should have a dataset with enough data and various classes to make our proposed methodology more challenging. Historical documents can be a suitable example of this family of datasets. Recently, preserving and analyzing historical documents in digital format has become an important task for historians. Most of these documents are degraded and have complex and/or unusual structures, requiring novel or custom detection methods. Detecting typographical features of interest to historians not only provides researchers with more targeted searching capability when accessing collections. It can also facilitate the large-scale study of bibliographic or textual changes over time [3]. In this paper, first we focus on information tables, a visual feature that appears frequently in historical scientific print documents. Then we propose a method that performs bounding box extraction via weakly supervision.

Historians of science have long emphasized the importance of visual evidence for the communication of scientific proof. Tables provided a new way for scientists to structure information in digestible ways, aiding the speeed and reliability of communicating information as well as establishing new protocols of correlational thinking. While we think of tables as natural to any scientific communication today, this process of the typographic organization of knowledge was both varied and historically complex. The ability to detect and study tables at large scale can help us better understand the origins and evolution of this key form of scientific communication.

Identifying tables in historical scientific print involves addressing some unique challenges. Tables are commonly structured as matrices of cells with a row-column structure, but there is some variation in their layout; this is particularly true of tables in historical print. Additionally, tables appear amidst a number of other visual conventions that share typographic features. These can include diagrams, maps illustrations, and ornamental items such as headpieces.

In developing a novel approach to table detection that addresses these challenge areas, we build on a number of recent advancements in machine learning-based approaches [4] to image detection. The advent of deep convolutional neural networks has generated new possibilities for object detection tasks, including table detection in documents. We used a CNN-based model for table detection, making several customisations. To boost reliability and accuracy, we trained our model with a dataset that consisted of multiple classes in addition to the table class, because the number of pages with tables in our historical dataset is relatively small. Our general model also had to be able to account for multiple table formats, including tables with and without ruling lines.

For example, in the [5], they used Faster Recurrent Convolutional Neural Network (Faster-RCNN) [2] as their table detector algorithm and train and evaluate their results on UNLV dataset [6]. The UNLV dataset contains only 427 document images with tables, which raises questions regarding its generalizability. If we use the proposed model by [5] on dataset which consist of different classes we will get a considerable number of false detected table (FP) because the model trained just based on table images.

The main contribution of this paper is proposing a method that produces enough labeled data and extracts bounding boxes from a limited available ground truth in hand. This way of training overcomes the problem of manual labeling and extracting bounding boxes of big datasets. To this end, we propose three phases of training. First, we train on a small subset that contains the ground-truth bounding box labels. Then, in the second phase, we apply our weakly supervision bounding box extraction that is using false positive detected samples from phase. Finally, we use the trained model from phase 2 to produce more (pseudo) labels and perform bounding box extraction for table detection on a much larger dataset.

This article is organized as follows: Section 2 surveys existing work in the field of table detection. Section 3 outline of proposed methodology, notably the technique of "Weakly supervised bounding box extraction" and prepossessing. Section 4 describes our experiments with this method and our results; Section 5 summarizes our conclusions to date and describes next steps.

2 Related work

There is a considerable body of existing research on table detection. Most approaches to this task cannot accommodate the identification of tables with different structures or formats. In other words, generalization is the most challenging problem in table detection.

We can divide existing research on table detection into two categories: the first one depends on deep learning methods and the second is based on layout analysis. [7] uses components and structural features of tables to detect them. [8] proposed one of the first methods for analysing tables, the T-Recs recognition system. This method detects tables by extracting word bounding boxes and segmenting them by bottom-up clustering. T-Recs depends heavily on word bounding boxes. [9] detects tables by recognizing column and row line separators and classifying them by SVM classifier. This method is suitable only for tables with ruling lines. [10] proposed one of the first method for spotting tables in PDF files by detecting and merging multi-lines with more than one text segment. However, this method cannot be used for multi-column documents. [11] localize tables in PDF files by using visual separators and geometric content layout information. It detects tables with or without ruling lines. [12] recognize tables based on computing an optimal partitioning of a document into some number of tables. This approach does not depend on ruling lines, headers, etc. Rather, the table quality measures is defined based on the correlation of white space and vertical connected component analysis. [13] used Hidden Markov Models (HMMs) in a different way to analyse document images. This approach computes feature vectors of white spaces between texts by using HMMs.

[14] extracts the size of text blocks and then compares their heights with the average height in the page. This approach can detect tables based on predefined rules. [15] finds regions as tables based on intersection between horizontal and vertical lines. Although this is an effective method, it entirely depends on the presence of ruling lines and a specific table structure.

Recent research has produced promising results in table detection by improving applications of deep learning algorithms specially in computer vision tasks. One well known algorithm in detection tasks is Faster-RCNN [2]. [16] uses a data-driven system that does not need heuristics to detect tables and their structures. This method detects columns, rows, and cell positions by carrying out Faster-RCNN on on the ICDAR 2013 dataset [17]. Gilani [5] uses Faster-RCNN on the UNLV dataset [6]. While this method is good for detecting tables with different structures, it is limited by its small training sample set, which contains only 427 table images. It also lacks some of the image classes pertinent to our historical dataset, such as ornamental items and diagrams, and so cannot be generalized as our model.

3 Proposed Method and Dataset

Building on existing approachs, we optimized our model with samples of false detection, as well as true detection, and measures to reduce the cost of labeling and extracting boxes. We used Faster-RCNN [2] as our detection algorithm, which is the third generation of RCNN [18] based models. RCNN are "Rich feature hierarchies for accurate object detection and semantic segmentation" that use a selective search to propose regions of interest that can be classified by a Convolutional Neural Network (CNN). This process is very time-consuming; to improve effciency of RCNN based algorithms, Girshick introduced the Fast-RCNN [19], which decreases the computational costs of detection tasks by implementing the Region of Interest Pooling technique. In the end, by replacing Region Proposal Network (RPN) instead of selective search in Faster-RCNN, Girschick introduced the ultimate version of RCNN based model. Faster-RCNN is a well-known and exemplary model for detecting natural scene images; hence we need to prepare our dataset before implementing it. In Faster-RCNN, a feature map is extracted from a pre-trained model based on natural scene images.

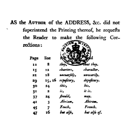
For training Faster-RCNN, we required labelled data and bounding boxes for each typographical element. The most prevalent datasets that are applied for table detection tasks are as follows:

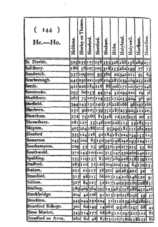
- Marmot [20]: contains 2000 Chinese and English documents. It also consists of different page layouts like one-column and two columns as well as various types of tables;
- UNLV [6]: contains 427 document images in different subjects such as newspapers, business letters etc.;
- ICDAR 2013 [17]: contains 128 documents that were used for analysing document images. It involves different domains of table competition tasks like table structure recognition, table location and their combination.

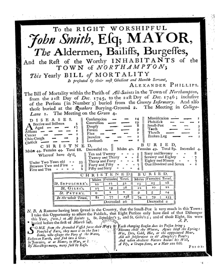
As mentioned in Section 1, our historical dataset contains a variety of table structures and visual objects similar to tables. To date, no table detection tasks have been undertaken with respect to historical documents. To evaluate our model, we used two new datasets that are drawn from collections of historical documents relevant to the history of scientific knowledge:

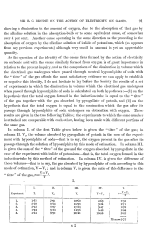
– ECCO: contains 32 million document images drawn from the Eighteenth-Century Collections Online database, which consist of over 200,000 documents published in the British Isles in the eighteenth century (INSERT CITATION). This dataset represents documents from a number of different genres printed over the course of one century in a single geographic region.
– NAS: contains 2 million document images drawn from the proceedings of five national academy of sciences (France, Sweden, Russia, Germany, and Britain) from 1666 to 1916. This resource provides documents from multiple languages and time periods that all belong to a related textual domain of scientific communication.

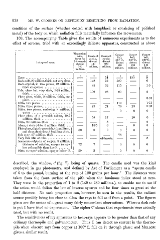
Because these datasets contain documents with different typographic conventions from different time periods and multiple kinds of visual representation, they pose significantly greater challenges than previous datasets. We present some examples in Fig. 1.


(a)


(b)


(c)


(d)


(e)

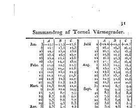

(f)

Fig. 1: a, b and c are samples of table documents from ECCO. d, e and f are samples of table documents from NAS.

3.1 Preprocessing

As mentioned above, the pre-trained model in Faster-RCNN is based on natural scene images. To make the images in our dataset comprehensible for Faster-RCNN, we transformed them to more closely resemble natural images. We used two well-known image processing applications for transforming images, Distance transform [21] and the Gabor filter [1]. The first method, distance transformation, gives us a blurred version of the initial image, changing intensities in the foreground so that they are more easily distinguished from the background. The transformed images can aid object recognition by facilitating machine intuition about the place of white background, text, and typographical elements. In [5], the author compute the distance between background and foreground by merging three different types of distance transforms such as Euclidean, Linear and Max distance transform. Fig. 2 represents the distance transformed of d, e and f images from Fig. 1.

(a) (b) (c)

Fig. 2: Results of Distance transformed on images d, c and f from Fig. 1.

Like distance transformation, the second method we used to convert document image files to natural images, the Gabor filter [1], separates contents from background to make an image more comprehensible to a machine. The Gabor filter is used to analyze the texture of images. Fig. 3 displays some images transformed with the Gabor filter. It is a sinusoidal signal that detects frequency of images in a particular direction. Extracting image features involves convolving a set of Gabor filters with different directions to the image. This filter consist of imaginary and real parts which are in orthogonal orientations. Equation(1) is demonstrating Gabor filter function and its variables:

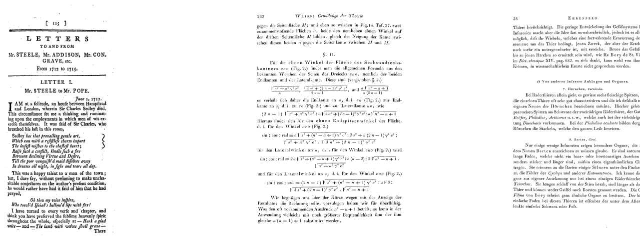
$$g(x,y;\lambda,\theta,\psi,\sigma,\gamma) = e^{-\frac{x'^2+\gamma^2 y'^2}{2\sigma^2}+2\pi i\frac{x'}{\lambda}+\psi i} \tag{1}$$

where λ =wavelength, θ =orientation of the normal, σ =standard deviation of the Gaussian, ψ =phase offset, γ =spatial aspect ratio, $x' = xcos\theta + ysin\theta$, $y' = -xsin\theta + ycos\theta$ [1].

(a) (b) (c)

Fig. 3: Results of Gabor filter on images d, c and f from Fig. 1.

Training approach Our approach to training was unusual in that it used two classes. Most work in table detection uses relatively small datasets and trains only on one class. However, historical documents such as those in our dataset contain a wide array of image objects and a very unbalanced distribution of each classs. Because of the unbalanced distribution of each class, we deemed it more useful to train Faster-RCNN based on a no-table class (i.e. pages without any tables) as well as the table class. Many images in our dataset belong to the non-table category (see Fig. 4 for examples). The use of two classes prevents our model from overfitting on recognizing tables, and, consequently, decreasing the false positive rate.



(a) (b) (c)

Fig. 4: Images without any table (no-table).

To develop a reliable model for predicting tables in new images, we needed to train our model with an adequately large set of labeled data with bounding boxes. To acquire this data manually was not ideal, given the high cost involved. We developed an approach that would mitigate manual labour.
The difficulty of generating training sets for our model can be divided into two parts: extracting bounding boxes, and labelling for table and no-table classes. To automate bounding box extraction, we developed a new procedure that we call

"Weakly supervised bounding box extraction". This innovative technique makes productive use of false positives which represented by the algorithm 1 1.

Algorithm 1 The weakly supervised bounding box extraction model

```
//Inputs:
table_images ← list of images that contain at least one table
table_bboxes ← dictionary that maps each image to its corresponding bounding boxes

notable_images ← list of images that contain no table

//Pre-Processing:
for image in table_images do
    image ← Gabor_Preprocess(image)
end for
for image in notable_images do
    image ← Gabor_Preprocess(image)
end for

//Train on images with tables:
model ← FasterRCNN.train(
        table_images,
        {table_bboxes, "TABLE"})

//Extract bounding boxes for images without tables:
notable_bboxes ← model.evaluate(
        images = notable_images)

//Train on images with tables and notables:
model ← FasterRCNN.train(
        table_images ∪ notable_images,
        {table_bboxes, "TABLE"} ∪
        {notable_bboxes, "NOTABLE"})

//Output weakly supervised trained checkpoint:
return model
```

We began with two mini-subsets of table and no-table items that lacked bounding boxes. We had these two mini-subsets labeled manually using Amazon Mechanical Turk. Three workers classified each document image with either the table or no-table label depending on whether the image contained any table items. Labels assigned by at least 2/3 workers to a given image became ground truth data. This process gave us the necessary training dataset comprised of images from two classes with their labels and bounding boxes.

At the first step, we manually extracted bounding boxes around some table images. Then we trained our model based on the table mini-subset. The weight was based merely on tables. Then we applied the theory of weakly supervised

bounding box extraction, which supposes that if we train our model based on Label 1 and we get a test on images which belong to Label 2, we will expect to receive a lot of samples which have been wrongly detected as Label 1 (i.e. a high rate of false positive). Because the model is just trained on one class, it is overfitted to detect the defined class. But here we optimized the first proposed model by using detected bounding boxes around an undefined class and changing their wrongly detected labels to their actual labels. In this experiment, the second class acquires the label of no-table; it contains text and any typographical elements that do not belong to the table class. The proportion of no-table class items is much greater than the table class in historical scientific books and documents. As a result, automatically extracting bounding boxes around the undefined, no-table class decreases the costs of bounding box extraction. These extracted bounding boxes are not very accurate but they are relatively acceptable boxes around text (no-table class). It is because of this relative extraction that we call the technique weakly supervised bounding box extraction.
This process gave us the necessary training dataset comprised of images from two classes with their labels and bounding boxes.
In the next phase of training, we associated the table and no-table labels with bounding boxes using the Pseudo Labeling technique, a semi-supervised learning approach. Pseudo Labeling consists of three main steps which are demonstrated in Fig. 5.

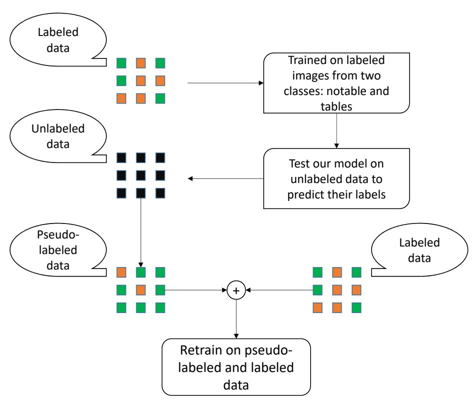


Fig. 5: pseudo-labeling approach pipeline.

The first phase of Pseudo Labeling uses the weakly supervised bounding box extraction technique already described. The second involves giving table and non-table images to our model to predict them. We called the resulting subset "Pseudo-Labelled Data." In the third phase, we gathered labelled data from Phase 1 and pseudo-labelled data from Phase 2 in one subset that we used to train our model.

4 Experiment And Results

In this section, after introducing the NAS and ECCO datasets, we first compare different image transformations techniques. Then, using the processed data, we apply our three-phase training approach and report the table detection results on these datasets.

– **NAS**:
  Our model training used images from the NAS subset altered using the Gabor filter and distance transformation. Then we compared the results of these two models on a mini-subset containing 500 distance transformed and 500 Gabor images. The proportion of the table class is 20 percent and the no-table is 80 percent. The following table 1 represent the results:

Table 1: comparison results of table detection with Faster-RCNN on mini-subset with 500 distance transformed and 500 Gabor images of NAS dataset.

transformation	precision	recall
Distance transform	76.77	85.03
Gabor filter	88.21	92.63

  We extracted bounding boxes around 922 table images and trained our algorithm based on them (Model A). We then tested Model A with 922 no-tables. As we expected, the result is overfitted on detecting tables; our Model (A) detected 879 tables from 922 no-tables, and there were a lot of false positive (FP) samples. To increase the accuracy, we decided to define two labels for table and no-table, determining that all images that do not contain tables would fall in the no-table class. We used weakly supervised bounding box extraction to put boxes around texts from 879 false positive tables and 43 true positive notables. Then we re-assigned false positive table items to the no-table class, creating a new mini-subset for a new training process. This mini-subset has 922 tables and 922 no-tables. In order to resolve the overfitting problem, we trained the algorithm based on this two labeled mini-subset (Model B). The next step was to generalize the model. We used pseudo labeling method to have more labeled images; after testing Model B on 1600 random images, we receive 576 tables and 1024 no-tables, calling this the NAS Final Subset. Then we trained the algorithm with the final subset (Model C). The steps of training, from A to C, are demonstrated in the table 2. At the end, we tested our model (model C) on bigger subset of NAS which consist of 2705 tables and 11607 no-tables. Table 3 represents the test results of this subset on our proposed model (model C) compared to the Faster-RCNN base model (model A) which has been trained on the "table" class without utilizing the weakly supervision bounding box extraction. It can be observed detected tables with our proposed idea in Fig. 6.

Table 2: Different steps of training our algorithm with different subsets of NAS.

Model	tables	notables
Model A	922	0
Model B	922	922
Model C	1498	1946

Table 3: Results of model C and A on the final subset of NAS

Method	Precision	Recall
Faster-RCNN(2 labels + weakly supervision bbox	81.19	86.44
Faster-RCNN(1 label)	54.2	93.66)

– **ECCO**:
For the ECCO dataset, as for the NAS data , we trained our model with the Gabor filter and distance-transformed images. Then we compared the results of these two models on mini-subset containing 500 distance transformed and 500 Gabor images. The proportion of table is 20 percent and no-table is 80 percent. The results are represented in table 4.

Table 4: comparison results of table detection with Faster-RCNN on mini-subset with 500 distance transformed and 500 Gabor images of ECCO dataset.

transformation	precision	recall
Distance transform	71.77	80.03
Gabor filter	82.21	87.63

In the first step, we extracted 785 bounding boxes around table images and trained our algorithm based on them (model A). In the second step, we tested Model A with 785 no-tables. As we expected, the result is overfitted on detecting tables; our model (A) detected 768 tables from 785 no-tables

(a) (b) (c)

Fig. 6: True detected tables on images d, c and f from Fig. 1.

and included false positive samples. As before, we used the weakly supervised bounding box extraction technique to obtain bounding boxes around 768 false positive tables and 17 true positive no-tables. Then, we changed the labels of these false positive tables to no-table and created a new mini-subset for training. This mini-subset has 785 tables and 785 no-tables. In order to resolve the overfitting problem, we trained the algorithm based on the two labeled mini-subsets (model B). To generalize the model, we used the pseudo labeling method to produce more labeled images. After testing Model B on 1124 random images, we were left with 437 tables and 687 no-table items. To further generalize the dataset for training, we added 437 tables and 687 no-tables to our mini-subset, calling this the ECCO Final Subset. The steps of training from A to C are demonstrated in the table 5. Finally, we tested our model (model C) on a bigger subset of NAS which consisted of 1230 tables and 5289 notables. Results of testing this subset on our proposed model (model C), versus the Faster-RCNN base model (model A) trained solely on "table" class without utilizing weakly supervision bounding box extraction, are represented in the table 6.

Table 5: Different steps of training our algorithm with different subsets of ECCO.

Model	tables	notables
Model A	785	0
Model B	785	785
Model C	1222	1472

Table 6: Results of model C and A on the final subset of ECCO

Method	Precision	Recall
Faster-RCNN(2 labels + weakly supervision bbox	77.15	71.46)
Faster-RCNN(1 label)	47.36	88.79)

From the comparison between our proposed method (model C) and the Faster-RCNN base model (model A) in tables 3 6, it can be observed that for model A, the precision is very low because of the high rate of false positive. Therefore, adding another label (no-table) along with using a weakly supervised bounding box extraction help to significantly improve the precision of the detection algorithm. It can also be observed from table 3 and table 6 that the precision and recall from the ECCO dataset is lower than that of the NAS dataset. The main reason is the particularly wide range of table structures in the ECCO data comprared to NAS.

5 Conclusion

There are not many comprehensive and challenging datasets in the domain of object detection tasks for the use case of table detection in document images. To address this issue, first, we introduced two new datasets. These datasets are more comprehensive and challenging since they contain various structures of tables as well as a different type of classes. In this article, we used Faster-RCNN as the state-of-the-art detection algorithm. The essential requirement for training on Faster-RCNN is to have enough labeled data, i.e., bounding boxes around typographical elements. To overcome the problem of shortage of labeled data on the proposed new datasets, we used our Weakly supervised bounding box extraction based on Faster-RCNN that by having a limited number of labeled images of one class, introduces a 3-step training procedure to produce labeled data and bounding boxes around their typographical elements. This model generalizes against different classes and accommodates various table structures. The innovative concept of Weakly supervised bounding box extraction technique used in this procedure, is particularly useful for limiting the manual cost of tedious bounding box labeling process. Future work on this project will involve making NAS dataset accessible to the public.

Acknowledgment

The authors thank the NSERC Discovery held by Prof. Cheriet for their financial support.

References

1. Tai Sing Lee. Image representation using 2d gabor wavelets. *IEEE Transactions on pattern analysis and machine intelligence*, 18(10):959–971, 1996.
2. Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pages 91–99, 2015.
3. Andrew Piper, Chad Wellmon, and Mohamed Cheriet. The page image: Towards a visual history of digital documents. *Book History*, 23(1):365–397, 2020.
4. Donald Michie, David J Spiegelhalter, CC Taylor, et al. Machine learning. *Neural and Statistical Classification*, 13(1994):1–298, 1994.
5. Azka Gilani, Shah Rukh Qasim, Imran Malik, and Faisal Shafait. Table detection using deep learning. In *2017 14th IAPR international conference on document analysis and recognition (ICDAR)*, volume 1, pages 771–776. IEEE, 2017.
6. Asif Shahab, Faisal Shafait, Thomas Kieninger, and Andreas Dengel. An open approach towards the benchmarking of table structure recognition systems. In *Proceedings of the 9th IAPR International Workshop on Document Analysis Systems*, pages 113–120, 2010.
7. Pallavi Pyreddy and W Bruce Croft. Tintin: A system for retrieval in text tables. In *Proceedings of the second ACM international conference on Digital libraries*, pages 193–200, 1997.

8. Thomas Kieninger and Andreas Dengel. Applying the t-recs table recognition system to the business letter domain. In *Proceedings of Sixth International Conference on Document Analysis and Recognition*, pages 518–522. IEEE, 2001.
9. Thotreingam Kasar, Philippine Barlas, Sébastien Adam, Clément Chatelain, and Thierry Paquet. Learning to detect tables in scanned document images using line information. In *2013 12th International Conference on Document Analysis and Recognition*, pages 1185–1189. IEEE, 2013.
10. Burcu Yildiz, Katharina Kaiser, and Silvia Miksch. pdf2table: A method to extract table information from pdf files. In *IICAI*, pages 1773–1785, 2005.
11. Jing Fang, Liangcai Gao, Kun Bai, Ruiheng Qiu, Xin Tao, and Zhi Tang. A table detection method for multipage pdf documents via visual seperators and tabular structures. In *2011 International Conference on Document Analysis and Recognition*, pages 779–783. IEEE, 2011.
12. Jianying Hu, Ramanujan S Kashi, Daniel P Lopresti, and Gordon Wilfong. Medium-independent table detection. In *Document Recognition and Retrieval VII*, volume 3967, pages 291–302. International Society for Optics and Photonics, 1999.
13. Ana Costa e Silva. Learning rich hidden markov models in document analysis: Table location. In *2009 10th International Conference on Document Analysis and Recognition*, pages 843–847. IEEE, 2009.
14. Dieu Ni Tran, Tuan Anh Tran, Aran Oh, Soo Hyung Kim, and In Seop Na. Table detection from document image using vertical arrangement of text blocks. *International Journal of Contents*, 11(4):77–85, 2015.
15. Basilios Gatos, Dimitrios Danatsas, Ioannis Pratikakis, and Stavros J Perantonis. Automatic table detection in document images. In *International Conference on Pattern Recognition and Image Analysis*, pages 609–618. Springer, 2005.
16. Sebastian Schreiber, Stefan Agne, Ivo Wolf, Andreas Dengel, and Sheraz Ahmed. Deepdesrt: Deep learning for detection and structure recognition of tables in document images. In *2017 14th IAPR international conference on document analysis and recognition (ICDAR)*, volume 1, pages 1162–1167. IEEE, 2017.
17. Max Göbel, Tamir Hassan, Ermelinda Oro, and Giorgio Orsi. Icdar 2013 table competition. In *2013 12th International Conference on Document Analysis and Recognition*, pages 1449–1453. IEEE, 2013.
18. Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587, 2014.
19. Ross Girshick. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 1440–1448, 2015.
20. Jing Fang, Xin Tao, Zhi Tang, Ruiheng Qiu, and Ying Liu. Dataset, groundtruth and performance metrics for table detection evaluation. In *2012 10th IAPR International Workshop on Document Analysis Systems*, pages 445–449. IEEE, 2012.
21. Heinz Breu, Joseph Gil, David Kirkpatrick, and Michael Werman. Linear time euclidean distance transform algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(5):529–533, 1995.