Semi-supervised Single-Shot Object Detection for Table Detection in Scanned Documents

Jamiu Bello, Xishuang Dong, and Xiangfang Li(✉)

CREDIT Center, Prairie View A&M University, Prairie View, USA
{jbello,xidong,xili}@pvamu.edu

Abstract. Table detection played a vital role in scanned document mining and understanding. The complex nature of tables has made table detection cumbersome and resource-intensive. In this paper, a novel two-path semi-supervised single-shot object detection framework was proposed for automatic detection of table in scanned documents. The proposed framework includes a shared VGG16 network and a two-path single-shot object detection model comprising of supervised and unsupervised subnetworks for table detection and classification. CascadeTabNet General dataset was employed to validate the effectiveness of the proposed framework. Experimental results demonstrated that the model implemented under the proposed framework is robust across various amount of available annotated documents (labeled data) for training. It can detect and classify tables in scanned document effectively even when training on very limited labeled data. For example, the precision of the proposed model is within 3% of a supervised model (SSD300) when training on only 10% of labeled data and 90% of unlabeled data.

Keywords: Table Detection · Semi-supervised Learning · Object Detection · Scanned Document

1 Introduction

In recent years, a lot of the traditional paper-based transactions are shifted online and being automated [17]. This business change has resulted in the rapid growth of the digital version of documents and has necessitated digitization of documents like paper-based invoices and receipts by scanning them. Table offers important information and provides structure for information in the documents. In many cases, documents contain various types of table-based information that is presented in various forms and layout [6]. Specifically, table can either be open or closed, where the open table has no borderlines while the closed table uses borderlines to separate rows and columns [26].

Because of the importance of tables in data representation and information delivery, table detection has become increasingly important and it has attracted a lot of attention. Table detection in scanned documents aims at detecting key information from tables [3]. Several methods for table detection were developed including rule-based approach [7, 10] and deep learning-based methods [1, 6] for closed table detection [11] and open table detection [18]. In addition, the use of statistical learning has been explored [25] to

K. Arai (Ed.): FTC 2023, LNNS 815, pp. 430–441, 2023.
https://doi.org/10.1007/978-3-031-47457-6_28

detect and recognize tables in scanned documents. So far deep learning-based methods have recorded the best performance based on the robust and excellent complex feature extraction techniques, achieving state-of-the-art result in table detection. However, most of the deep learning approaches are supervised learning-based approaches, which require a huge amount of labeled data to train a model successfully. It is well known that annotation of large amount of scanned documents for table detection is very expensive and resource intensive. Moreover, it could not be achieved in a real-time fashion. On the other hand, there exist large amount of unannotated documents (unlabeled data).

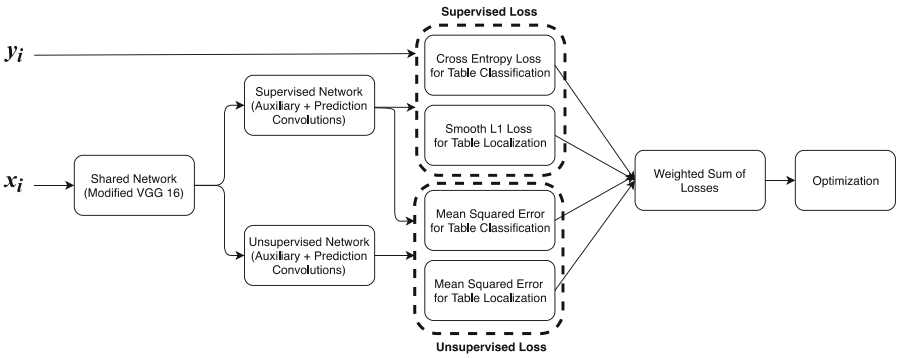

Fig. 1. Framework of the proposed two-path semi-supervised learning (TPSSL). Sample x_i is the input. Label y_i is available only for the labeled inputs (annotated documents). The associated cross-entropy loss is to evaluate the table classification loss and the smooth L_1 loss is to evaluate the table localization loss for the supervised subnetwork. Mean squared error is employed to evaluate both the table classification and table localization loss of the unsupervised subnetwork.

In this paper, we proposed a semi-supervised single-shot detection framework to detect tables in scanned documents. The proposed approach uses very limited annotated documents while taking advantage of large amount of unannotated documents for training. It is a one-stage table detection framework built upon two-path replica subnetwork and single-shot detection (SSD) technique [14]. The detailed framework is shown in Fig. 1. Specifically, it consists of a shared VGG network for low-level feature extraction, one supervised subnetwork and one un-supervised subnetwork. The supervised subnetwork calculates the loss from the labeled data using cross entropy and smooth L_1 loss for classification and localization losses, respectively. The unsupervised subnetwork made use of mean squared error to determine the classification and localization loss. The overall loss is calculated from the weighted sum of both the supervised and the unsupervised losses which is further used to optimize the proposed model. The proposed framework was validated on table detection with CascadeTabNet General dataset including ICDAR 2019 and other public datasets. The experimental results demonstrate the effectiveness of the proposed method even with very limited amount of labeled training data.

In summary, the contributions of this paper include:

– A novel two-path semi-supervised object detection framework is proposed for table detection and classification in scanned documents. The framework is comprised

of supervised and unsupervised subnetwork built upon single-shot object detection method to detect and classify tables.
– The proposed framework made use of both labeled and unlabeled data to successfully localize and classify tables within scanned document.
– Experimental results demonstrated that the model implemented under the proposed framework is robust across various amount of available labeled training data.
– The proposed method can detect tables in scanned document effectively even when training on very limited labeled data. For example, the precision of the proposed model is within 3% of a supervised model (SSD300) when training on only 10% of labeled data and 90% of unlabeled data.

The rest of the paper is organized as follows. Section 2 reviews some related work. The proposed framework is explained in detail in Sect. 3. Experimental results and analysis are presented in Sect. 4. Section 5 concludes the paper.

2 Related Work

The growing use of digital documents in the last few decades resulted in the increasing trend seen in information extraction and conversion of a paper-based document to digital format. The introduction and wide use of portable document format (PDF) contributed to this rise in demand. Online commerce and education have also made digital form of information exchange a must at this time. This increasing demand in digital records necessitated the focus on document understanding and automation [23]. Table detection from documents has brought a tremendous value in the field of commerce and education [23].

As technology advances, table detection problem evolved into machine learning problems resolved by support vector machine (SVM) [12], sequence labeling [22] and ensemble [2]. Cesarini et al. [2] developed Tabfinder that first converted the document into an MXY tree representation and then searched for blocks surrounded by lines. Perez et al. [16] proposed layout heuristic with k-nearest neighbor for recognizing table structure. With further advancement in machine learning, other robust machine learning methods were also proposed. Farrukh et al. [4] modeled a reasoning approach that combined k-means clustering, random forest, and markov logical network for table structure recognition. Rashid et al. [19] developed a multilayer neural network for table structure recognition through using some artificial features for cell classification and post-process to enhance the result.

The outstanding success recorded by deep neural networks inspired researchers to explore deep learning for table detection and localization in documents. Gilani et al. [6] first modeled table detection as a time-series learning problem using Faster RCNN. Arif and Shafait [1] made use of semantic color-coding to improve the performance of Faster R-CNN. He et al. [9] developed page segmentation using FCN. Hao et al. [8] proposed table detection in portable document formats using deep neural networks. Recently, Siddiqui et al. [21] modeled DeCNT by integrating Faster RCNN with deformable CNN. This framework adopted its receptive field-based to its input and achieved excellent performance on IC-DAR2017. Most of deep neural network methods have generated state-of-the-art performance in table detection and structure recognition. They have also

tackled the earlier problem of format and layout challenges encountered in the rule-based detection approach.

In summary, most of the current deep learning-based methods are based on supervised learning. It requires a large amount of labeled data to implement the detection task. However, annotating tables and table cells has a huge cost implication. Semi-supervised learning is a technique that is able to use both labeled data and unlabeled data to reduce the efforts of data annotation for building machine learning models. Regarding this benefit, we propose a novel framework of deep semi-supervised learning for table detection.

3 Methodology

The proposed framework is for table detection to localize and classify tables present on scanned images of documents. Suppose the training data consists of a total N input images, from which M are labeled. The input x_i ($i \in 1 \ldots N$) is represented as a set of the scanned images of documents. S represents the set of labeled inputs, $|S| = M$. For each $i \in S$, there exist a corresponding label $y_i \in 1 \ldots C$, where C shows the number of classes available in the dataset. The framework of the model proposed in this paper is shown in Fig. 1. The proposed framework of this model and corresponding learning procedures is to evaluate the network for each training input x_i with the supervised path and the unsupervised path to complete two tasks. The first task is to learn how to detect tables using labeled data while the second task is to optimize the learning algorithm for table detection without the ground truth.

As shown in Fig. 1, for each of the training input x_i sample i.e., the scanned images and the corresponding class label y_i and object coordinates passes through the supervised branch of the network. The proposed framework built upon the single-shot object detection (SSD) [14] passes the input through a shared network for low level feature extraction. This shared network layer, derived from VGG16 backbone, comprises of seven CNN blocks responsible for extracting low-level features from the input image. It generated feature maps as its output to the auxiliary layer. The auxiliary convolution blocks, which are additional CNN blocks integrated with the shared network, further extracts features at multiple scales and progressively decreases the size of the input at each subsequent layer. The various sizes of the generated feature maps from various convolution filters are then passed to the last layer, the prediction layer, where the table predictions are made. The output of the prediction layer is anchor boxes and class labels for the detected tables. This predicted anchor boxes and labels are then matched with the class labels and bounding boxes of tables. The overlap between the anchor boxes and the ground truth of table localization, referred to as intersection-over-union (IoU) are measured. The predictions that have IoU greater than 0.5 in value are considered as positive samples, while the lower IoU results are considered as negative samples or wrong predictions. All redundant anchor boxes are removed through a post-processing concept known as non-maximum suppression.

The learning process is initiated with positive and negative training samples obtained from the matched anchor boxes. The main objective is to start the training with positive samples and gradually train the model to shift the predictions towards the ground truth. Each grid cell with a positive sample generates a probability vector, and a four elements

vector to adjust the position of the anchor box to match the ground truth box. After every training step, the priors, in which the overall cost function diminishes, are kept and re-adjusted towards the ground-truth boxes better. The model convergence is achieved when the loss curve becomes stable or when the deviation between the ground-truth and prior is close to zero.

Furthermore, the proposed semi-supervised framework aimed to optimize the learning procedure using unlabeled data. The unlabeled training input xi passes through the shared network and then through the supervised and unsupervised network. Two prediction vectors which are new representations of the input data are generated from the supervised and the unsupervised networks, respectively. The supervised and the unsupervised network are similar but the output from them is stochastic in nature. It implies that there will be difference between the two prediction vectors for the same input sample. Given that the original input x_i is the same, this difference can be seen as an error and thus minimizing the mean square error (MSE) is a reasonable objective in the learning procedure. The overall network is evaluated for the classification and localization loss with the supervised path including share network and supervised network and the unsupervised path including shared network and unsupervised network, which results in prediction vectors c_i^{sup} and l_i^{sup} for table classification and localization in the ii supervised branch and c_i^{unsup} and l_i^{unsup} for table classification and localization ii in the unsupervised branch. The two pair of vectors are used to evaluate the overall loss which is given by:

$$L = -\frac{1}{|B|}\sum_{i\in B\cap S}\left(f_{softmax}\left(c_i^{sup}\right)\left[y_i\right] + f_{smoothL1}\left(l_i^{sup}\right)\left[y_i\right]\right) + w(t)\times\frac{1}{K|B|}\left(\sum_{i\in B}\left\|c_i^{sup} - c_i^{unsup}\right\|^2 + \left\|l_i^{sup} - l_i^{unsup}\right\|^2\right) \quad (1)$$

B is the mini-batch within the learning process. The overall objective loss L is a weighted sum of the localization (L) and confidence (C) losses for both the supervised and unsupervised branches. This loss consists of two main components. The first component is used to evaluate the supervised loss just for the labeled inputs with cross entropy loss and smooth L1 loss. The second unsupervised component evaluates for all inputs (both labeled and unlabeled), and then penalizes the variant predictions for the same training input xi by calculating the Mean Squared Error (MSE). To enable the model combine the result of the supervised loss and unsupervised loss, the latter is scaled by a weighting function w(t) that is time-dependent, starting from zero; it ramps up through a Gaussian curve.

4 Results and Analysis

4.1 Datasets

We employed the Cascade TabNet General dataset [17] which was created by combining various public datasets published for table detection including ICDAR 2019 [5], Marmot [13] and Github. Specifically, ICDAR 2019 dataset includes images of word and latex documents with text. Marmot dataset is a publicly available dataset published by the Institute of Computer Science and Technology of Peking University. It contains texts from Chinese and English. The last part of the dataset was extracted from Github repositories including 1,934 scanned document with 2,835 tables.

4.2 Experimental Settings

In this experiment, our proposed model is employed to implement table detection. The key hyper-parameters of the proposed model are: Batch size: 4, Number of epoch: 100, Optimizer: SGD, Learning rate: 0.0001.

4.3 Evaluation Metrics

We evaluate the performance of our model using different evaluation metrics [15], which include average Precision, average Recall, and average Fscore. In addition, Mean Average Precision (mAP) was applied for evaluation where Average Precision (P) is calculated for every class from the area under the precision-recall curve. Precision measures the percentage of the model predictions that are correct after matching them with the ground truth. And Recall (R) measures the rate of all the possible ground-truth boxes being detected. The below equations show the formulae for Precision and Recall.

$$\mathrm{P} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{2}$$

$$\mathrm{R} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{3}$$

$$\mathrm{Fscore} = \frac{2 \times \mathrm{P} \times \mathrm{R}}{\mathrm{P} + \mathrm{R}} \tag{4}$$

True Positive (TP) is the number of predictions by the model that has intercession over union (IoU) greater than 0.5 with ground-truth boxes. False Positive (FP) is the number of predictions by the model with the IoU less than 0.5 with ground-truth boxes. False Negative is the number of ground-truth boxes that the trained model does not detect. Their confidence score ranks predictions from highest to lowest. Finally, average Precision (AP) is evaluated as the average of maximum precision values at the chosen recall values. Mean Average Precision (mAP) is simply the average of APs over all the classes as shown in Eq. 5. The more significant the mAP, the higher performance is the model achieved.

$$mAP = \frac{1}{C} \sum\nolimits_{c \in C} AP_C \tag{5}$$

4.4 Experimental Results

We explore different evaluation metrics to measure the performance of the proposed model. Table 1 showed the performance comparison between the proposed model and two state-of-the-art supervised learning-based models including Single-shot detection (SSD) [14] and YOLOV3 [20]. Compared to the baselines, the proposed models can achieve competitive performance even learning on limited labeled training data. For example, the proposed model (TPSSL) can achieve similar precision and competitive F-score and recall when training on 40% labeled data. Moreover, in general, the performance of TPSSL was improved when training with more labeled training data. It

Table 1. Comparing performance between baselines and proposed models (TPSSL). The baselines including SSD300, SSD512, and YOLOV3 were built through training on fully labeled data while we applied 10%, 20%, 30%, 40% and 50% labeled data together with unlabeled data to accomplish learning of the proposed models.

Model	Precision	Recall	F-score	mAP
SSD300 [14]	97.86	42.73	59.49	81.16
SSD512 [14]	96.30	44.53	60.90	84.35
YOLOV3 [20]	49.48	65.42	56.35	64.96
TPSSL (10%)	94.81	33.39	49.82	61.16
TPSSL 20%)	94.65	37.48	53.70	65.91
TPSSL (30%)	93.50	40.27	56.29	71.89
TPSSL (40%)	97.67	40.57	57.32	76.27
TPSSL (50%)	97.11	40.79	57.45	74.88

indicated that more labeled training data will help enhance performance for the proposed model, which is consistent with observations of previous work on semi-supervised learning [24].

Table 2 presented the performance comparison with different ratios of labeled training data combined with various IoU. It can be observed that with the increased IoU, the performance was dropped significantly since higher IoU required more overlaps between predicted bounding-box and ground truth to obtain correct predictions for calculating prediction, recall, and F-score. In addition, with less labeled training data, the trend of performance reduction with increased IoU is more significantly. For example, comparing cases between 10% and 50% available labeled training data, with increased IoU, the precision is decreased from 94.81% to 10.78% for the 10% case while it is reduced from 97.11% to 27.20% for 50% case. It is expected that more labeled training data will build a more robust model, which is also confirmed through comparing baselines and TPSSL.

Moreover, Figs. 2, 3, and 4 explore the effects of different learning rates on the performance of the proposed model. It is observed that the proposed model performed successfully across all three learning rates. It is also observed that the model performance improved slightly with decreased learning rate. For example, the precision, recall, and F-score all improved slightly when there are 40% labeled training data with decreased learning rate. This can be attributed to the fact that the model learns slowly in smaller steps with lower learning rates, thereby is able to converge with higher performance, unlike a larger learning rate seems to make the proposed model converge too quickly and result in lower performance. We also observed that the bars dropped dramatically with the increased IoU when the labeled training data is very limited (say, 10%), while the bars dropped much slower with the increased IoU when more labeled training data are available (say, 40%). This implies that performances with higher IoU are greatly impacted with fewer labeled training samples.

Table 2. Comparing performances generated with different ratios of labeled training data under various IoU.

Baseline 1: SSD300			
IoU	Precision (%)	Recall (%)	F-score (%)
0.5	97.86	42.73	59.48
0.6	97.78	42.67	59.41
0.7	97.71	42.61	59.34
0.8	95.72	41.54	57.94
0.9	78.49	33.85	47.30
Baseline 2: YOLOV3			
IoU	Precision (%)	Recall (%)	F-score (%)
0.5	49.48	65.42	56.35
0.6	45.30	59.52	51.45
0.7	34.64	43.79	38.68
0.8	20.47	24.67	22.38
0.9	5.57	6.34	5.92
TPSSL (10% Labeled Data)			
IoU	Precision (%)	Recall (%)	F-score (%)
0.5	94.81	33.39	49.82
0.6	94.80	31.76	47.58
0.7	87.13	27.95	42.32
0.8	47.22	14.16	21.78
0.9	10.78	3.09	4.80
TPSSL (30% Labeled Data)			
IoU	Precision (%)	Recall (%)	F-score (%)
0.5	93.50	40.27	56.29
0.6	93.00	39.91	55.85
0.7	89.40	37.86	53.19
0.8	76.89	32.06	45.25
0.9	25.27	9.11	13.39

(*continued*)

Table 2. (*continued*)

TPSSL (50% Labeled Data)			
IoU	Precision (%)	Recall (%)	F-score (%)
0.5	97.11	40.79	57.45
0.6	96.34	40.29	56.82
0.7	94.61	39.30	55.53
0.8	78.43	32.40	45.85
0.9	27.20	10.48	15.13

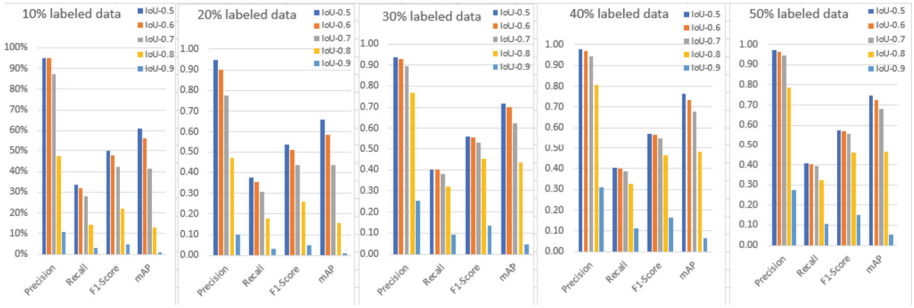

Fig. 2. Different performances generated with learning rate 0.001 different ratios of labeled data, namely 10%, 20%, 30%, 40% and 50% x-axis is for different evaluation metrics while y-axis is for performance. Different color bars illustrate different batch IoU, where dark blue bars are for 0.5, brown bars are for 0.6, grey bars are for 0.7, Orange bars are for 0.8 and light blue bars are for 0.9. (Colour figure online)

In summary, increasing the amount of labeled data for training improves the performance of the proposed model. The model recorded an acceptable performance even with very little amount of labeled data, which shows that the proposed model is robust on detecting tables from scanned documents. Meanwhile, to achieve optimal performance, the learning rate needs to be selected carefully.

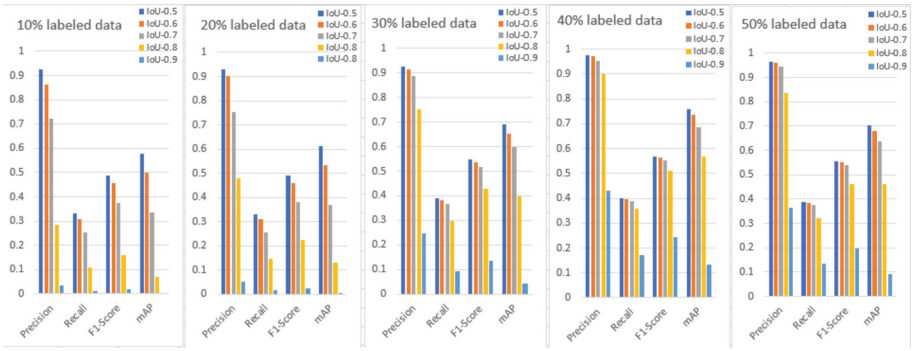

Fig. 3. Different performances generated with learning rate 0.0001 different ratios of labeled data, namely 10%, 20%, 30%, 40% and 50% x-axis is for different evaluation metrics while y-axis is for performance. Different color bars illustrate different batch IoU, where dark blue bars are for 0.5, brown bars are for 0.6, grey bars are for 0.7, Orange bars are for 0.8 and light blue bars are for 0.9. (Colour figure online)

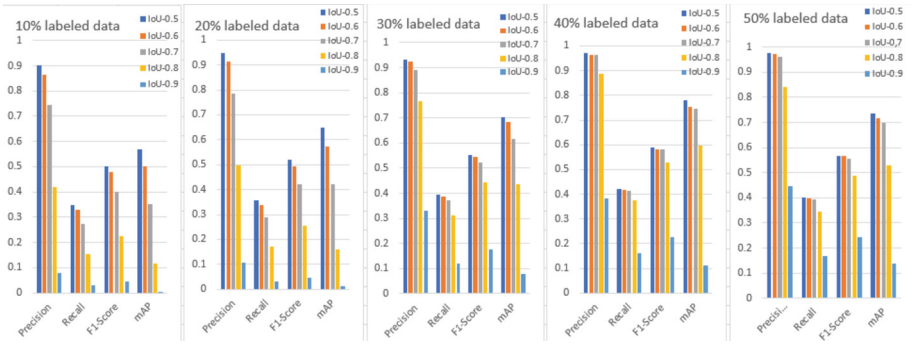

Fig. 4. Different performances generated with learning rate 0.00005 different ratios of labeled data, namely 10%, 20%, 30%, 40% and 50% x-axis is for different evaluation metrics while y-axis is for performance. Different color bars illustrate different batch IoU, where dark blue bars are for 0.5, brown bars are for 0.6, grey bars are for 0.7, Orange bars are for 0.8 and light blue bars are for 0.9. (Colour figure online)

5 Conclusion

In this paper, a novel semi-supervised table detection and classification framework is proposed for scanned documents. The framework is comprised of supervised and unsupervised subnetwork built upon single-shot object detection method. It made use of both labeled and unlabeled data to successfully localize and classify tables within scanned document. Experimental results demonstrated that the proposed method can detect tables in scanned document effectively even when training on very limited labeled data. For example, the precision of the proposed model is within 3% of a supervised model (SSD300) when training on only 10% of labeled data and 90% of unlabeled data. This is extremely valuable in practice because annotation of large amount of scanned documents

for table detection is very expensive and resource intensive. Moreover, it could not be achieved in a real-time fashion. On the other hand, there exist large amount of unannotated documents (unlabeled data). Thus, the proposed method is an attractive approach for table detection in scanned documents when there are very limited resources and/or time for data annotations which happens very often in reality.

Acknowledgment. This research work is supported by NSF award 2205891 and NASA under award number 80NSSC22KM0052. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of NSF or NASA.

References

1. Arif, S., Shafait, F.: Table detection in document images using foreground and background features. In: 2018 Digital Image Computing: Techniques and Applications (DICTA), pp. 1–8. IEEE (2018)
2. Cesarini, F., Marinai, S., Sarti, L., Soda, G.: Trainable table location in document images. In: 2002 International Conference on Pattern Recognition, vol. 3, pp. 236– 240. IEEE (2002)
3. Embley, D.W., Hurst, M., Lopresti, D., Nagy, G.: Table-processing paradigms: a research survey. IJDAR **8**(2), 66–86 (2006)
4. Farrukh, W., et al.: Interpreting data from scanned tables. In: 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), vol. 2, pp. 5–6. IEEE (2017)
5. Gao, L., et al.: ICDAR 2019 competition on table detection and recognition (cTDAR). In: 2019 International Conference on Document Analysis and Recognition (ICDAR), pp. 1510–1515. IEEE (2019)
6. Gilani, A., Qasim, S.R., Malik, I., Shafait, F.: Table detection using deep learning. In: 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), vol. 1, pp. 771–776. IEEE (2017)
7. Green, E., Krishnamoorthy, M.: Recognition of tables using table grammars. In: Proceedings of the Fourth Annual Symposium on Document Analysis and Information Retrieval, pp. 261–278 (1995)
8. Hao, L., Gao, L., Yi, X., Tang, Z.: 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)
9. He, D., et al.: Multi-scale FCN with cascaded instance aware segmentation for arbitrary oriented word spotting in the wild. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3519–3528 (2017)
10. Hu, J., Kashi, R.S., Lopresti, D.P., Wilfong, G.: Medium-independent table detection. In: Document Recognition and Retrieval VII, vol. 3967, pp. 291–302. SPIE (1999)
11. Itonori, K.: Table structure recognition based on textblock arrangement and ruled line position. In: Proceedings of 2nd International Conference on Document Analysis and Recognition (ICDAR 1993), pp. 765–768. IEEE (1993)
12. Kasar, T., Barlas, P., Adam, S., Chatelain, C., Paquet, T.: Learning to detect tables in scanned document images using line information. In: 2013 12th International Conference on Document Analysis and Recognition, pp. 1185–1189. IEEE (2013)
13. Li, M., Cui, L., Huang, S., Wei, F., Zhou, M., Li, Z.: Tablebank: a benchmark dataset for table detection and recognition. arXiv preprint arXiv:1903.01949 (2019)

14. Liu, W., et al.: SSD: single shot multibox detector. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9905, pp. 21–37. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46448-0_2
15. Padilla, R., Netto, S.L., Da Silva, E.A.: A survey on performance metrics for object-detection algorithms. In: 2020 International Conference on Systems, Signals and Image Processing (IWSSIP), pp. 237–242. IEEE (2020)
16. Perez-Arriaga, M.O., Estrada, T., Abad-Mota, S.: TAO: system for table detection and extraction from PDF documents. In: The Twenty-Ninth International Flairs Conference (2016)
17. Prasad, D., Gadpal, A., Kapadni, K., Visave, M., Sultanpure, K.: CascadeTabNet: an approach for end to end table detection and structure recognition from image-based documents. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 572–573 (2020)
18. Raja, S., Mondal, A., Jawahar, C.V.: Table structure recognition using top-down and bottom-up cues. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.M. (eds.) ECCV 2020. LNCS, vol. 12373, pp. 70–86. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58604-1_5
19. Rashid, S.F., Akmal, A., Adnan, M., Aslam, A.A., Dengel, A.: Table recognition in heterogeneous documents using machine learning. In: 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), vol. 1, pp. 777–782. IEEE (2017)
20. Redmon, J., Farhadi, A.: YOLOv3: an incremental improvement. arXiv preprint arXiv:1804.02767 (2018)
21. Siddiqui, S.A., Malik, M.I., Agne, S., Dengel, A., Ahmed, S.: DeCNT: deep deformable CNN for table detection. IEEE Access **6**, 74151–74161 (2018)
22. e Silva, A.C.: Learning rich hidden Markov models in document analysis: table location. In: 2009 10th International Conference on Document Analysis and Recognition, pp. 843–847. IEEE (2009)
23. Subramani, N., Matton, A., Greaves, M., Lam, A.: A survey of deep learning approaches for OCR and document understanding. arXiv preprint arXiv:2011.13534 (2020)
24. Van Engelen, J.E., Hoos, H.H.: A survey on semi-supervised learning. Mach. Learn. **109**(2), 373–440 (2020)
25. Wangt, Y., Phillipst, I.T., Haralick, R.: Automatic table ground truth generation and a background-analysis-based table structure extraction method. In: Proceedings of Sixth International Conference on Document Analysis and Recognition, pp. 528–532. IEEE (2001)
26. Zanibbi, R., Blostein, D., Cordy, J.R.: A survey of table recognition. Doc. Anal. Recognit. **7**(1), 1–16 (2004)