Deep Layout Analysis of Multi-lingual and Composite Documents

Takwa Ben Aïcha Gader and Afef Kacem Echi(✉)

National Superior School of Engineering, University of Tunis, LR: LATICE, Tunis, Tunisia
takoua.benaicha@enis.tn, ff.kacem@gmail.com

Abstract. It is crucial to accurately analyze the layout to convert document images to high-quality text. With the emergence of publicly available, large ground-truth datasets, deep-learning models have demonstrated their effectiveness in detecting and segmenting document layouts. This study presents a deep learning technique for document structure analysis, an important stage in the optical character recognition (OCR) system. Our method employs the YOLOv7 (Only Look Once version 7) model, a highly efficient and precise object detection model trained on the DocLayNet database. The trained YOLOv7 model quickly and efficiently identified and categorized different document components, such as caption, list item, text, table, section header, and picture. Regarding accuracy and efficiency, our evaluation demonstrates that the suggested method beats existing strategies, with strong generalization ability for diverse document layouts, text styles, and scripts.

Keywords: Document Layout Analysis · Object Detection · Deep Learning · Optical Character Recognition · Text Detection · YOLOV7

1 Introduction

The optical character recognition (OCR) method translates document images into machine-readable text. It has become increasingly important daily, allowing us to digitize and store documents quickly, search and index text, and extract information from scanned images or documents. OCR has many practical applications, such as document management, data entry, and text analysis. One of its primary benefits is its ability to save time and reduce the need for manual data entry. Thus it allows the preservation and easy access to essential documents. This can be useful in various domains, such as healthcare, finance, and legal industries, where a vast amount of data must be processed and stored. The quality of scanned documents and the presence of noise or distractions in images can impact the accuracy of OCR systems.

OCR technology faces a significant obstacle in layout analysis, which entails identifying and differentiating various sections of a document, such as text, images, tables, and graphics. This is important because it allows the OCR system

A. Bennour et al. (Eds.): ISPR 2023, CCIS 1940, pp. 76–89, 2024.
https://doi.org/10.1007/978-3-031-46335-8_7

to accurately extract the text from the document and ignore non-text elements. However, layout analysis can be challenging due to the variations in document layouts and image noise or distractions. For example, a document may contain multiple columns, different font sizes and styles, or overlapping text and graphics. These variations can make it difficult for an OCR system to identify and extract the text accurately. One solution for this issue is utilizing deep learning models to examine the document's arrangement and precisely recognize the written content. These models were used for image segmentation and other tasks like text detection and text recognition, giving challenging results. However, developing an effective layout analysis system remains an active area of research, as many challenges remain to be addressed.

This study proposes a method utilizing the powerful object detection model YOLOv7 for analyzing the structure of OCR documents and detecting text through deep learning techniques. The YOLOv7 is used to accurately and efficiently detect the scanned documents objects and to classify them into one of the six classes; picture, caption, list item, table, section header, and text. YOLOv7 is a powerful deep-learning model for object detection in images and videos proposed in 2022. A convolutional Neural Network (CNN) can be trained to locate and identify objects in an image or video by predicting bounding boxes around them and classifying them into given classes. This version of the YOLO model is one of the most recent updates, renowned for its speedy and precise object detection abilities. Real-time object detection and classification make it a valuable solution for a range of applications, including self-driving cars, security systems, and facial recognition. One of the critical features of YOLOv7 is its ability to make predictions on full images in a single forward pass rather than requiring a sliding window approach or region proposal method. This makes it faster and more efficient than some other object detection models.

The YOLOv7 model was trained using the DocLayNet database [6] in this work, and the results of our studies show that it beats existing techniques in terms of accuracy and efficiency. We tested it on Latin and Arabic documents of various layouts and text styles, and it shows good generalization ability. The proposed YOLOv7-based approach significantly improves OCR page layout analysis and text detection.

In the following sections, we will explore the proposed method in more detail. Firstly, we will examine related works in Sect. 2. Then, in Sect. 3, we will introduce the YOLOV7 model proposed for page layout analysis. Afterward, some experimental results are discussed. In Sect. 5, some conclusions and prospects are finally given.

2 State of the Art

Document layout analysis is an essential step in OCR systems, which aims to break down documents into different components such as text, headings, tables, figures, lists, and more. This process, known as document segmentation, helps structure the document's content and facilitates the transcription of the text by the OCR system. By separating the text from other document elements, the OCR system can focus on transcribing the text without being disturbed by formatting and layout features. Additionally, accurate and efficient segmentation of documents into their different components can improve the overall structure and readability of the transcribed text. Therefore, document segmentation is necessary to ensure the success of OCR systems.

One of the previous works aimed at solving the document analysis problem is the article proposed by [5] in 1993. The "Document Spectrum" (or "Docstrum"), a method of structural layout analysis based on unsupervised hierarchical clustering of page components, is described in this work. This method accurately measures the tilt, line spacing, and text block spacing, thus offering three significant advantages over many other methods: Independence from title angle, independence from varied text spacings, and the capacity to manage local regions with different text orientations in the same image are all desirable.

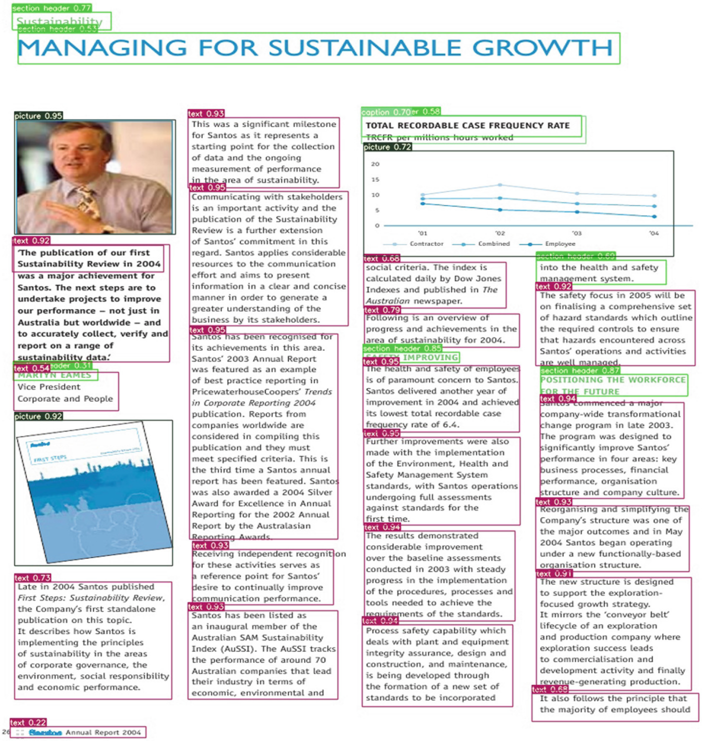
Several proposals have been made to solve the problem of layout analysis using deep learning methods. An example is the method proposed in 2012, a multiple-layer margin segmentation method for manuscripts with complex page layouts was presented in [1]. This technique employs straightforward and distinguishing characteristics obtained from the connected component level, resulting in the creation of resilient feature vectors. A multi-layer perceptron classifier is employed to classify connected components based on their pixel-level class. Following that, a voting process is used to refine the resulting segmentation and obtain the final categorization. The method was trained and tested on a dataset containing different complex margin layout formats, achieving a segmentation accuracy of 95%. In 2019, an innovative method was presented in [13] to identify logical structures in document images using visual and textual features. This technique uses two layers of LSTM recurrent neural networks to process the text in the identified zone, which is classified as a sequence of words, as well as the normalized position of each word concerning the page width and height. The labeled zones include abstract, title, author name, and affiliation. This approach achieved an overall accuracy of 96.21% on the publicly available MARG dataset. In 2022, a new deep learning approach [7] was presented to improve the understanding of document images. This system uses neural networks to convert physical documents into digital documents to extract the necessary information. The system relies on two autoencoder-decoder networks simultaneously segment text lines and non-textual components and identify non-textual elements. Tests conducted on RDCL2019 showed that this approach is more stable and adaptable to new formats than previous commercial and editorial systems. A recent study [11] proposed a new hybrid spatial attention network (HSCA-Net) to improve document layout analysis by enhancing feature extraction capability. The HSCA-Net comprises a spatial attention module, a channel attention module, and a lateral attention link. The performance of this network was evaluated on public datasets; PubLayNet, ICDAR-POD, and Article Regions.

In recent years, Object detection models like YOLO and SSD are useful in object recognition tasks; therefore, we decided to apply them to the problem of layout analysis in scanned texts. It should be noted that object detection involves determining the presence and location of objects in images or videos by answering two questions: where is the object located in the image, and what type of object is it? There are two primary categories in object detection: two-stage and one-stage approaches. In two-stage object detection, detection and recognition are separated, as with RCNN and Fast RCNN. In one-stage object detection, however, detection and recognition are merged into a single step, as with YOLO and Single-Shot-Detector (SSD). An example of a state-of-the-art method that uses object detection models for document analysis was presented in 2019 [2]. An innovative method based on YOLO was presented for table detection in scanned documents. This innovative approach includes adaptive adjustments, an anchor optimization strategy, and two post-processing methods, for example. The anchor optimization strategy uses k-means clustering to find more suitable table anchors. Additionally, post-processing techniques remove unnecessary white spaces and noisy page objects to improve the accuracy of table margins and UI scores. We also note the [4], which presented a deep neural network influenced by natural scene object detectors. The model was trained and evaluated using labeled samples from a large publicly available dataset, demonstrating the utility of object detectors in layout analysis. In 2022, a table detection model based on YOLO was proposed in [10]. The authors integrated an involution layer into the network backbone to improve its ability to learn the spatial layout features of tables. Additionally, they designed a simple pyramid network to enhance the model's efficiency, resulting in a 2.8% improvement in accuracy compared to YOLOv3 and a 1.25x increase in speed. Finally, we cite the work presented in [6], which introduced the DocLayNet database. This database contains 80863 pages that have been manually annotated from various sources, showcasing different layouts for the purpose of general document layout analysis. The authors trained multiple deep learning models, including YOLOV5, to highlight the diversity of the database.

We referenced the YOLOV7 as it is a one-stage object detector. It locates objects in an image by placing bounding boxes around them and identifies the objects' class or class probabilities, representing the object's name (as illustrated in Fig. 1). In this research, we use it to detect various document elements, such as text, titles, tables, figures, and lists. By training this model on the *DocLayNet* dataset of document images, we can automatically segment documents into their different components and extract meaningful text for further processing by the OCR system. This deep learning approach can significantly enhance the accuracy and efficiency of OCR systems.

3 Proposed Method

We present in this section the offered deep learning method for solving the issue of document layout analysis using the YOLOV7 model. This latter is a state-of-the-art object detection model well-suited for document layout analysis because it can accurately identify and classify different objects in images or videos using



Sustainability

MANAGING FOR SUSTAINABLE GROWTH

'The publication of our first Sustainability Review in 2004 was a major achievement for Santos. The next steps are to undertake projects to improve our performance – not just in Australia but worldwide – and to accurately collect, verify and report on a range of sustainability data.'

MARTYN EAMES
Vice President
Corporate and People

Late in 2004 Santos published *First Steps: Sustainability Review*, the Company's first standalone publication on this topic. It describes how Santos is implementing the principles of sustainability in the areas of corporate governance, the environment, social responsibility and economic performance.

This was a significant milestone for Santos as it represents a starting point for the collection of data and the ongoing measurement of performance in the area of sustainability.

Communicating with stakeholders is an important activity and the publication of the Sustainability Review is a further extension of Santos' commitment in this regard. Santos applies considerable resources to the communication effort and aims to present information in a clear and concise manner in order to generate a greater understanding of the business by its stakeholders.

Santos has been recognised for its achievements in this area. Santos' 2003 Annual Report was featured as an example of best practice reporting in PricewaterhouseCoopers' *Trends in Corporate Reporting 2004* publication. Reports from companies worldwide are considered in compiling this publication and they must meet specified criteria. This is the third time a Santos annual report has been featured. Santos was also awarded a 2004 Silver Award for Excellence in Annual Reporting for the 2002 Annual Report by the Australasian Reporting Awards.

Receiving independent recognition for these activities serves as a reference point for Santos' desire to continually improve communication performance.

Santos has been listed as an inaugural member of the Australian SAM Sustainability Index (AuSSI). The AuSSI tracks the performance of around 70 Australian companies that lead their industry in terms of economic, environmental and social criteria. The index is calculated daily by Dow Jones Indexes and published in *The Australian* newspaper.

Following is an overview of progress and achievements in the area of sustainability for 2004.

SAFETY IMPROVING

The health and safety of employees is of paramount concern to Santos. Santos delivered another year of improvement in 2004 and achieved its lowest total recordable case frequency rate of 6.4.

Further improvements were also made with the implementation of the Environment, Health and Safety Management System standards, with Santos operations undergoing full assessments against standards for the first time.

The results demonstrated considerable improvement over the baseline assessments conducted in 2003 with steady progress in the implementation of the procedures, processes and tools needed to achieve the requirements of the standards.

Process safety capability which deals with plant and equipment integrity assurance, design and construction, and maintenance, is being developed through the formation of a new set of standards to be incorporated

TOTAL RECORDABLE CASE FREQUENCY RATE
TRCFR per millions hours worked

into the health and safety management system.

The safety focus in 2005 will be on finalising a comprehensive set of hazard standards which outline the required controls to ensure that hazards encountered across Santos' operations and activities are well managed.

POSITIONING THE WORKFORCE FOR THE FUTURE

Santos commenced a major company-wide transformational change program in late 2003. The program was designed to significantly improve Santos' performance in four areas: key business processes, financial performance, organisation structure and company culture.

Reorganising and simplifying the Company's structure was one of the major outcomes and in May 2004 Santos began operating under a new functionally-based organisation structure.

The new structure is designed to support the exploration-focused growth strategy. It mirrors the 'conveyor belt' lifecycle of an exploration and production company where exploration success leads to commercialisation and development activity and finally revenue-generating production.

It also follows the principle that the majority of employees should

26 Santos Annual Report 2004

Fig. 1. Example of image with detected and recognized components.

a single convolutional neural network (CNN). The YOLOv7 object detection model, proposed by Alexey Bochkovskiy et al. in [8], is a real-time object detection tool that uses single-stage detection. It was released in July 2022 and has been recognized as an exceptionally efficient and accurate object detector. It outperforms other deep learning models and YOLO versions in speed and accuracy, with performance ranging from 5 FPS to 160 FPS [8]. In [8], the authors discuss the improvements made to YOLOv7 and how it has set a new benchmark for object detection performance. They also describe the optimization techniques and strategies implemented to enhance the training and inference processes.

3.1 The YOLOv7 Architecture

In this study, we employed the original YOLOv7 architecture, in which image frames pass through the backbone network for feature extraction. These features are then combined and processed further in the neck of the network and passed on to the head, where the model makes predictions about the presence, location, and class of different objects contained in the image and generates bounding boxes around them (see Fig. 3 for an illustration of the used architecture). In the following, we detail the different key components of the YOLOV7 architecture:

- **Backbone Network**: the backbone network of the YOLOv7 model is a get the training lead head and auxiliary head labels simultaneously feature from the input image. The backbone selection is a key step, as it will improve object detection performance. In the original YOLOv7 paper, the authors used the E-ELAN deep CNN architecture (see Fig. 3) as the backbone network. It employs group convolution to enhance the cardinality of the incorporated features and then combines the features of different groups through a shuffle-and-merge method. This approach improves the feature representation and increases the efficiency of computation and usage of parameters (Fig. 2).

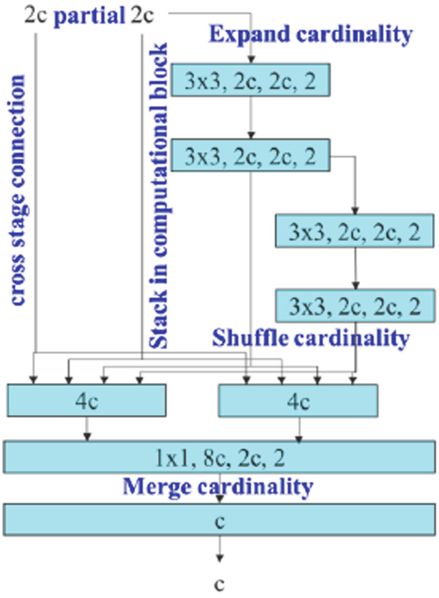


Fig. 2. The used extended ELAN (E-ELAN) [8].

- **Neck**: the neck of the YOLOV7 refers to the layers between the backbone network and the head layers (output layers). Its architecture is responsible for creating a feature pyramid, a set of feature maps at different scales that the head layers use to make predictions about the presence and location of objects in an image. The neck layers generally consist of convolutional, pooling, and up-sampling layers. For enhancement, authors of [8] used Attention Guided Feature Pyramid Network (FPN), feature refinement block (FRB), or Pyramid Attention Network (PAN).
- **Head**: In the context of the YOLOv7 architecture, the head refers to the output layers of the network that are responsible for making predictions about the presence, location, and type of objects within an image. The head layers take the features generated by the backbone and neck layers as input and use them to detect objects and produce the final bounding box coordinates

and class probabilities for each object detected in the image. The head layers are generally composed of a combination of convolutional layers and fully connected layers. We used the lead head guided label assigner as a head architecture. The training lead head and auxiliary head labels were obtained simultaneously through optimization using lead head prediction and ground truth. Otherwise, it generates the final predictions, and soft labels are derived from them. It's important to note that YOLOv7 has a Lead Head and an Auxiliary Head, which share the same loss function (see Fig. 3. Both heads are trained using the same soft labels. Still, the Lead Head is used to generate the final predictions, and the Auxiliary Head is used to refine the predictions of the Lead Head. This way, YOLOv7 can achieve high accuracy and real-time performance.

- **Anchors**: anchors in the YOLOv7 model are pre-defined bounding boxes that help the model identify objects in an image. These bounding boxes are chosen to have different sizes and shapes so that they can detect objects of various sizes and proportions. The model uses these anchor boxes as a starting point for object detection and then fine-tunes them to fit the objects in the image. These anchor boxes cover a range of object sizes and aspect ratios expected to be found in the training data.
- **Loss Function**: the YOLOv7 training process utilizes a composite loss function that includes multiple terms. The key components of this loss function include:
    - To enhance the model's ability to predict the exact location of objects in an image, a localization loss term is utilized. This term computes the mean squared error between the expected bounding box coordinates and the real ground-truth bounding box coordinates.
    - To improve its accuracy in predicting the class probabilities of objects in images, the model uses a confidence loss term. This term calculates the cross-entropy between the predicted class probabilities and the actual class labels.
    - To enhance the model's ability to anticipate the class labels of objects in an image, a classification loss term is computed by measuring the cross-entropy between the predicted and actual class labels.

  The YOLOv7 loss function is a weighted sum of the localization, confidence, and classification Losses calculated for all the anchor boxes with an object present in the ground truth. When training, the YOLOV7 model aims to minimize this loss function by adapting the model's parameters to decrease the differences between the predicted bounding box coordinates, class probabilities, and class labels with ground-truth values. So, the model learns to recognize objects, their location, and their classification in the image.

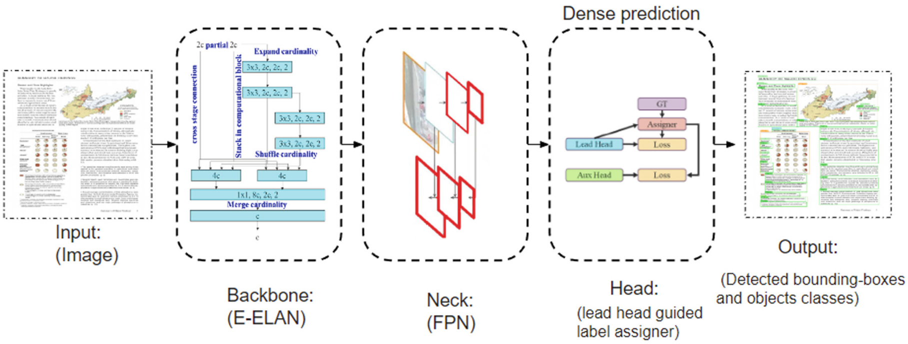


Fig. 3. The used YOLOV7 architecture.

4 Experimental Results

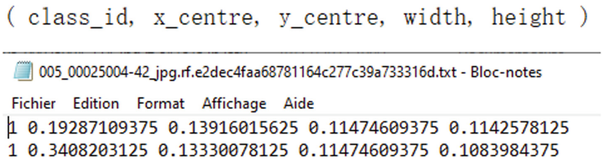
4.1 Used Database

We used the PubLaynet [12] database to test the YOLOV7 model for Document structure analysis. This dataset is utilized for analyzing document layout. It comprises of images of articles and research papers, along with annotations for diverse elements present on these pages, including text, lists, and figures. Over a million publicly available PDF articles from PubMed Central were mechanically matched with their XML representations to build the dataset [3]. In PubLayNet, there are several categories of document layout such as Caption, Page-header, Footnote, List-item, Page-footer, Formula, Picture, Table, Section-header, Text, and Title. However, we limit our training to six classes - caption, list item, picture, section header, table, and text - to enhance accuracy and precision.

Database Pre-processing. The database annotation is available in the COCO (JSON) format. The COCO format is a specific way of organizing labels and metadata for an image dataset using the JSON structure. However, to train our YOLOV7 model, the dataset must be in the YOLO format, so we reformated it in the acceptable format. In the YOLO format, each image in the dataset is associated with a single text file. If an image does not contain objects, it will not have a corresponding text file. The text file for each image contains rows of information, one row for each object presented in the image. For each row, we have five columns which indicate: Class-id denotes the class of the current object, x-center, and y-center represent the coordinate of the bounding box center, and a width and a height indicate the width and the height of the current bounding box (see Fig. 4 for an example).

To transform the database annotations from the COCO format to the YOLO format, we followed the following steps:

- Obtaining image-related information such as image_id, image_width, image_height, etc.
- Retrieve the annotations for a particular image using image_id.



```
( class_id, x_centre, y_centre, width, height )
```

005_00025004-42_jpg.rf.e2dec4faa68781164c277c39a733316d.txt - Bloc-notes
Fichier Edition Format Affichage Aide

```
1 0.19287109375 0.13916015625 0.11474609375 0.1142578125
1 0.3408203125 0.13330078125 0.11474609375 0.1083984375
```

Fig. 4. Example of annotations in the YOLO format.

- Open a text file for the current image at the output path.
- Retrieve the bounding box properties for each object in the image.
- Save the retrieved annotations for the current image in a text file.
- Upon processing all the annotations for the current image, close the associated text file.
- Repeat the above steps for all images in the dataset.

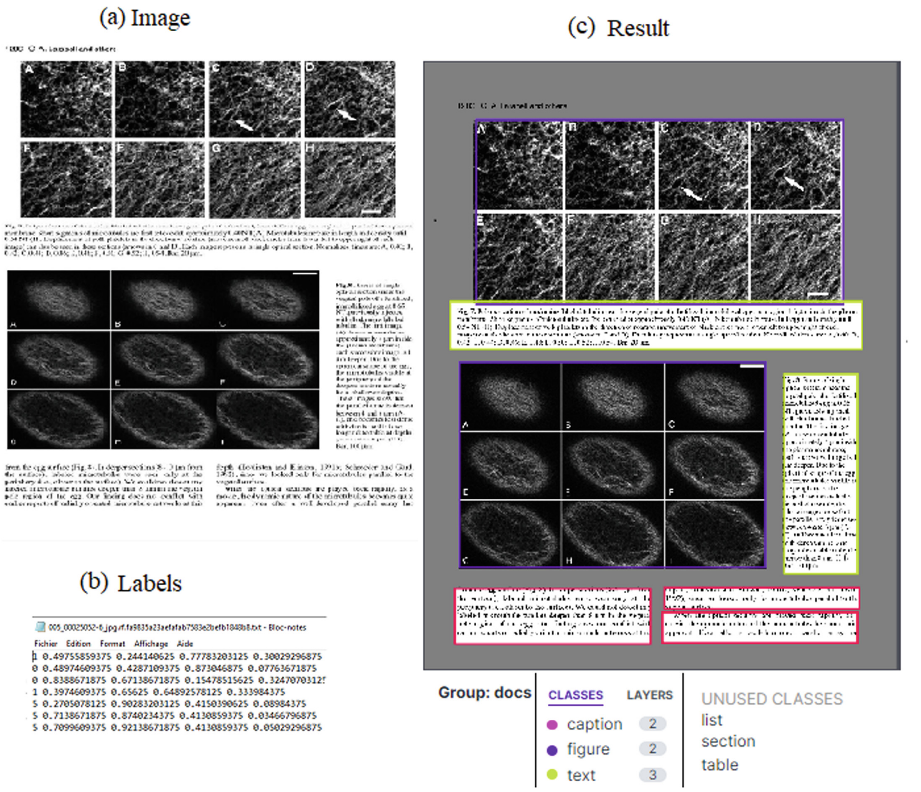
Before using the converted dataset to train the YOLOV7 model, it was essential to ensure that it met the requirements and was converted correctly. To do this, we used the Roboflow online tool and checked sample images from the dataset (an example is shown in the Fig. 5).

4.2 Training

Instead of training our model from scratch, we opted to utilize transfer learning by starting with a generic COCO pre-trained checkpoint downloaded from [9]. Transfer learning is a methodology where a model previously trained for one task is adapted and refined to suit a different yet related task. This approach allowed us to take advantage of the knowledge and features learned by the pre-trained model and apply them to our document layout analysis task, saving us both labeled data and computational resources. As a result, we achieved better performance on the task with less cost.

For training the YOLOV model, we used multi-scale training, which involves training the model using images of different sizes. This helps the model to learn to detect objects of various sizes and in different contexts, which improves its ability to generalize and make more accurate predictions. Multi-scale training aims to prepare the model to detect objects in real-world images, regardless of their size. This can be done by randomly resizing input images during training or with different image scales like 32×32, 416×416, etc. We split the database into 70% for the training set, 20% for the validation, and 10% for the test and trained the model on the training dataset for 500 epochs using the training settings listed in Table 1.

The performance of the Yolov7 model on the training and validation sets is presented in Fig. 6, which displays three different types of losses: box loss, target loss, and classification loss. Box loss measures the algorithm's ability to identify the object's center and accurately predict its bounding box. Target loss evaluates the probability that an object will be found in a suggested zone of



```
1 0.49755859375 0.244140625 0.77783203125 0.30029296875
0 0.48974609375 0.4287109375 0.873046875 0.07763671875
0 0.8388671875 0.67138671875 0.15478515625 0.32470703125
1 0.3974609375 0.65625 0.64892578125 0.333984375
5 0.2705078125 0.90283203125 0.4150390625 0.08984375
5 0.7138671875 0.8740234375 0.4130859375 0.03466796875
5 0.7099609375 0.92138671875 0.4130859375 0.05029296875
```

Fig. 5. Example of annotations in the YOLO format.

interest. Classification loss indicates the algorithm's capacity to classify an object correctly. The model's precision, recall, and average accuracy improved rapidly during training, as evidenced by the decreasing box, target, and classification loss values. The model improved precision, recall, and mAP after 200 epochs and achieved stability after 400.

4.3 Testing

Once the model was trained, it was utilized to analyze unseen images and video. Throughout the inference process, a confidence threshold of 0.1 was maintained. Figure 8 illustrates that the trained model could detect different objects presented on a document image and classify them to their accurate classes with prediction probabilities greater than 90% in most instances. We used the mAP (mean average precision) metric to evaluate the model's performance. An evaluation metric frequently utilized in object detection tasks. The mAP is calculated by calculating the Average Precision (AP) for each class and then averaging it over many classes (see Eq. 1). Precision (AP) is the proportion of true positive

Table 1. Training Parameters.

YOLOV7 model	hyperparameters
415 layers	$lr0 = 0.01$, $lrf = 0.1$, $momentum = 0.937$, $weight_{decay} = 0.0005$, $warmup_{epochs} = 3.0$, $warmup_{momentum} = 0.utilizeds_{lr} = 0.1$, $box = 0.05$, $cls = 0.3$, $cls_{pw} = 1.0$, $obj = 0.7$, $obj_{pw} = 1.0$, $iou_t = 0.2$, $anchor_t = 4.0$, $fl_{gamma} = 0.0$, $hsv_h = 0.015$, $hsv_s = 0.7$, $hsv_v = 0.4$, $degrees = 0.0$, $translate = 0.2$, $scale = 0.9$, $shear = 0.0$, $perspective = 0.0$, $flipud = 0.0$, $fliplr = 0.5$, $mosaic = 1.0$, $mixup = 0.15$, $copy_{paste} = 0.0$, $paste_{in} = 0.15$, $loss_{ota} = 1$

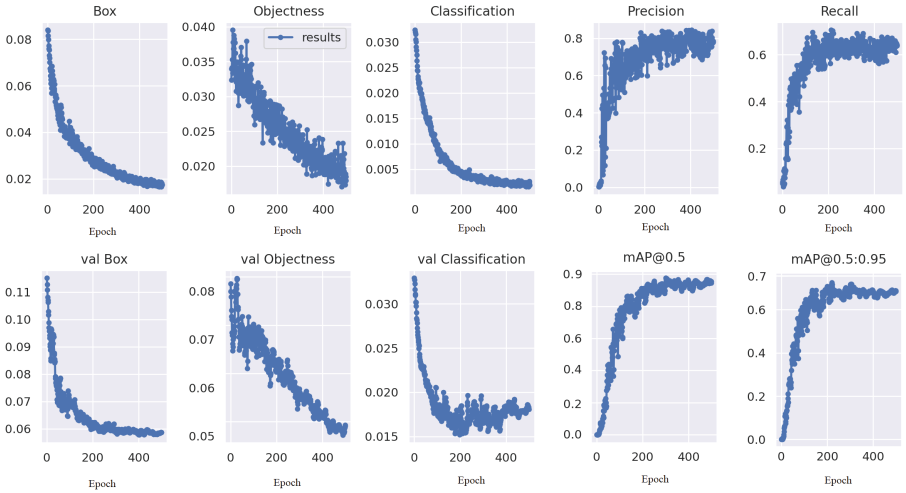


Fig. 6. The training and validation curves of box loss, objectness loss, classification loss, precision, recall, and mean average precision (mAP).

detections (correctly identified objects) among all positive detections (correct and incorrect).

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \qquad (1)$$

The mean average precision (mAP) offers a consolidated value that accurately reflects the overall performance of the model. It is widely used in computer vision competitions and research to compare the performance of different object detection models. Table 2 shows our results compared to Mask R-CNN, Faster R-CNN, and YOLOv5 models on the test database based on the mAP@0.5–0.95(and the Fig. 7 represents the precision curves of the six classes on the test dataset. The following formula calculates the precision:

$$mAP = \frac{TP}{TP + FP}; \tag{2}$$

where TP is the True positive and the FP is the False Positive.

The results show that the YOLOV7 model achieved the highest mAP compared to other models with respected gaps.

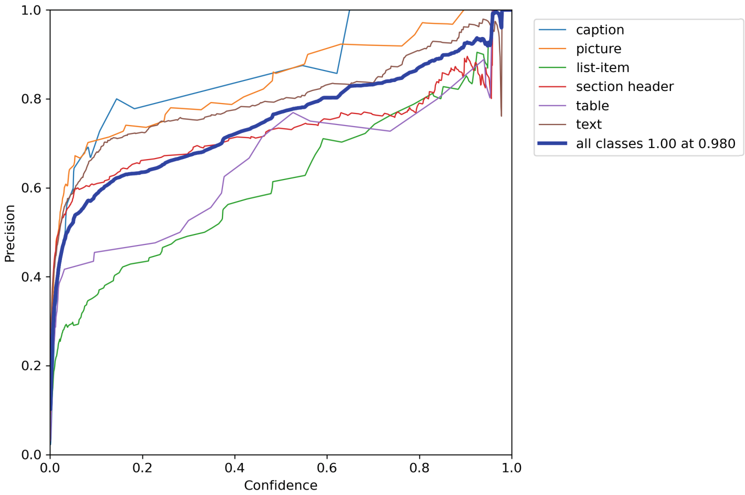


Fig. 7. Precision curves for the six classes for the test dataset.

Table 2. Prediction performance (mAP@0.5-0.95) of object detection models on DocLayNet test dataset.

Class	MRCNN [6]		FRCNN [6]	YOLO	
	R50	R101	R101	V5x6 [6]	**V7**
Caption	68.4	71.5	70.1	77.7	**80.9**
List item	81.2	80.8	81.0	86.2	**90.1**
Picture	71.7	72.7	72.0	77.1	**83.6**
Section-header	67.6	69.3	68.4	74.6	**84.0**
Table	82.2	82.9	82.2	86.3	**91.3**
Text	84.6	85.8	85.4	88.1	**92.7**
All	75,95	51,78	76,51	81,66	**87.1**

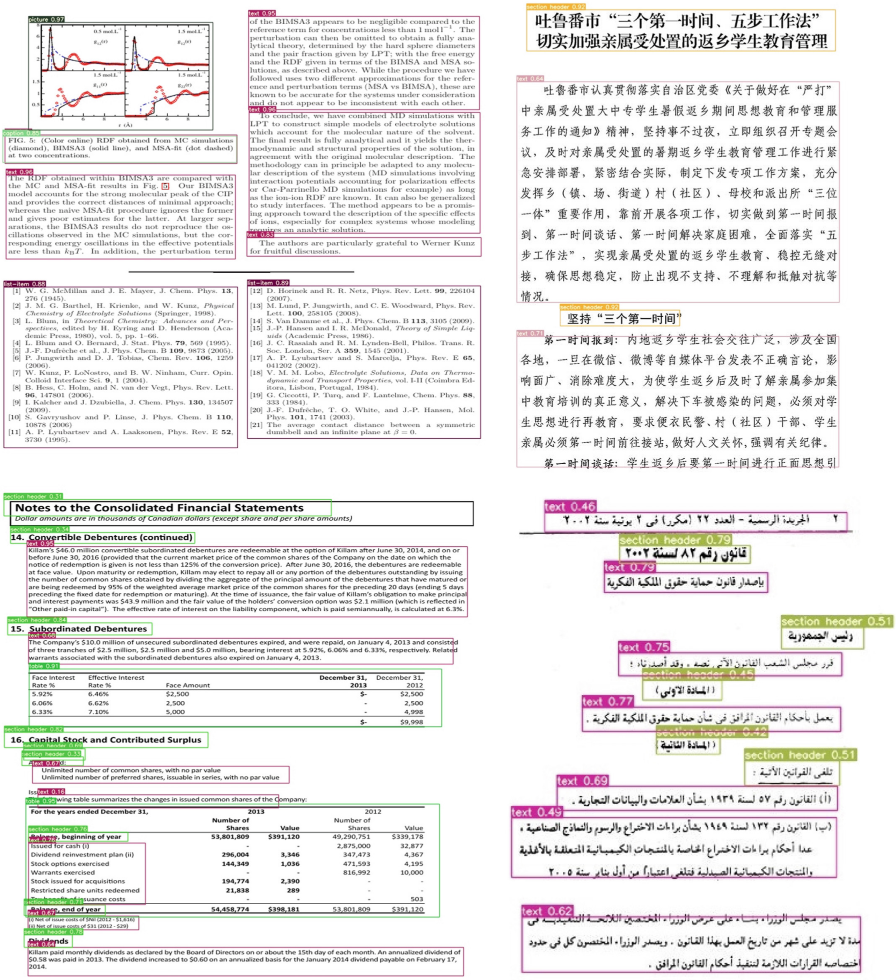


FIG. 5: (Color online) RDF obtained from MC simulations (diamond), BIMSA3 (solid line), and MSA-fit (dot dashed) at two concentrations.

The RDF obtained within BIMSA3 are compared with the MC and MSA-fit results in Fig. 5. Our BIMSA3 model accounts for the strong molecular peak of the CIP and provides the correct distances of minimal approach; whereas the naive MSA-fit procedure ignores the former and gives poor estimates for the latter. At larger separations, the BIMSA3 results do not reproduce the oscillations observed in the MC simulations, but the corresponding energy oscillations in the effective potentials are less than k_BT. In addition, the perturbation term of the BIMSA3 appears to be negligible compared to the reference term for concentrations less than 1 mol l^{-1}. The perturbation can then be omitted to obtain a fully analytical theory, determined by the hard sphere diameters and the pair fraction given by LPT; with the free energy and the RDF given in terms of the BIMSA and MSA solutions, as described above. While the procedure we have followed uses two different approximations for the reference and perturbation terms (MSA vs BIMSA), these are known to be accurate for the systems under consideration and do not appear to be inconsistent with each other.

To conclude, we have combined MD simulations with LPT to construct simple models of electrolyte solutions which account for the molecular nature of the solvent. The final result is fully analytical and it yields the thermodynamic and structural properties of the solution, in agreement with the original molecular description. The methodology can in principle be adapted to any molecular description of the system (MD simulations involving interaction potentials accounting for polarization effects or Car-Parrinello MD simulations for example) as long as the ion-ion RDF are known. It can also be generalized to study interfaces. The method appears to be a promising approach toward the description of the specific effects of ions, especially for complex systems whose modeling requires an analytic solution.

The authors are particularly grateful to Werner Kunz for fruitful discussions.

[1] W. G. McMillan and J. E. Mayer, J. Chem. Phys. **13**, 276 (1945).
[2] J. M. G. Barthel, H. Krienke, and W. Kunz, *Physical Chemistry of Electrolyte Solutions* (Springer, 1998).
[3] L. Blum, in *Theoretical Chemistry: Advances and Perspectives*, edited by H. Eyring and D. Henderson (Academic Press, 1980), vol. 5, pp. 1–66.
[4] L. Blum and O. Bernard, J. Stat. Phys. **79**, 569 (1995).
[5] J.-F. Dufrêche et al., J. Phys. Chem. B **109**, 9873 (2005).
[6] P. Jungwirth and D. J. Tobias, Chem. Rev. **106**, 1259 (2006).
[7] W. Kunz, P. LoNostro, and B. W. Ninham, Curr. Opin. Colloid Interface Sci. **9**, 1 (2004).
[8] B. Hess, C. Holm, and N. van der Vegt, Phys. Rev. Lett. **96**, 147801 (2006).
[9] I. Kalcher and J. Dzubiella, J. Chem. Phys. **130**, 134507 (2009).
[10] S. Gavryushov and P. Linse, J. Phys. Chem. B **110**, 10878 (2006)
[11] A. P. Lyubartsev and A. Laaksonen, Phys. Rev. E **52**, 3730 (1995).
[12] D. Horinek and R. R. Netz, Phys. Rev. Lett. **99**, 226104 (2007).
[13] M. Lund, P. Jungwirth, and C. E. Woodward, Phys. Rev. Lett. **100**, 258105 (2008).
[14] S. Van Damme et al., J. Phys. Chem. B **113**, 3105 (2009).
[15] J.-P. Hansen and I. R. McDonald, *Theory of Simple Liquids* (Academic Press, 1986).
[16] J. C. Rasaiah and R. M. Lynden-Bell, Philos. Trans. R. Soc. London, Ser. A **359**, 1545 (2001).
[17] A. P. Lyubartsev and S. Marcelja, Phys. Rev. E **65**, 041202 (2002).
[18] V. M. M. Lobo, *Electrolyte Solutions, Data on Thermodynamic and Transport Properties*, vol. I-II (Coimbra Editora, Lisbon, Portugal, 1984).
[19] G. Ciccotti, P. Turq, and F. Lantelme, Chem. Phys. **88**, 333 (1984).
[20] J.-F. Dufrêche, T. O. White, and J.-P. Hansen, Mol. Phys. **101**, 1741 (2003).
[21] The average contact distance between a symmetric dumbbell and an infinite plane at $\beta = 0$.

吐鲁番市“三个第一时间、五步工作法”
切实加强亲属受处置的返乡学生教育管理

吐鲁番市认真贯彻落实自治区党委《关于做好在“严打”中亲属受处置大中专学生暑假返乡期间思想教育和管理服务工作的通知》精神，坚持事不过夜，立即组织召开专题会议，及时对亲属受处置的暑期返乡学生教育管理工作进行紧急安排部署，紧密结合实际，制定下发专项工作方案，充分发挥乡（镇、场、街道）村（社区）、母校和派出所“三位一体”重要作用，靠前开展各项工作，切实做到第一时间报到、第一时间谈话、第一时间解决家庭困难，全面落实“五步工作法”，实现亲属受处置的返乡学生教育、稳控无缝对接，确保思想稳定，防止出现不支持、不理解和抵触对抗等情况。

坚持“三个第一时间”

第一时间报到：内地返乡学生社会交往广泛，涉及全国各地，一旦在微信、微博等自媒体平台发表不正确言论，影响面广、消除难度大，为使学生返乡后及时了解亲属参加集中教育培训的真正意义，解决下车被感染的问题，必须对学生思想进行再教育，要求便衣民警、村（社区）干部、学生亲属必须第一时间前往接站，做好人文关怀，强调有关纪律。

第一时间谈话：学生返乡后要第一时间进行正面思想引

Notes to the Consolidated Financial Statements
Dollar amounts are in thousands of Canadian dollars (except share and per share amounts)

14. Convertible Debentures (continued)

Killam's $46.0 million convertible subordinated debentures are redeemable at the option of Killam after June 30, 2014, and on or before June 30, 2016 (provided that the current market price of the common shares of the Company on the date on which the notice of redemption is given is not less than 125% of the conversion price). After June 30, 2016, the debentures are redeemable at face value. Upon maturity or redemption, Killam may elect to repay all or any portion of the debentures outstanding by issuing the number of common shares obtained by dividing the aggregate of the principal amount of the debentures that have matured or are being redeemed by 95% of the weighted average market price of the common shares for the preceding 20 days (ending 5 days preceding the fixed date for redemption or maturing). At the time of issuance, the fair value of Killam's obligation to make principal and interest payments was $43.9 million and the fair value of the holders' conversion option was $2.1 million (which is reflected in "Other paid-in capital"). The effective rate of interest on the liability component, which is paid semiannually, is calculated at 6.3%.

15. Subordinated Debentures

The Company's $10.0 million of unsecured subordinated debentures expired, and were repaid, on January 4, 2013 and consisted of three tranches of $2.5 million, $2.5 million and $5.0 million, bearing interest at 5.92%, 6.06% and 6.33%, respectively. Related warrants associated with the subordinated debentures also expired on January 4, 2013.

Face Interest Rate %	Effective Interest Rate %	Face Amount	**December 31, 2013**	December 31, 2012
5.92%	6.46%	$2,500	$-	$2,500
6.06%	6.62%	2,500	-	2,500
6.33%	7.10%	5,000	-	4,998
			$-	$9,998

16. Capital Stock and Contributed Surplus

A d:

Unlimited number of common shares, with no par value
Unlimited number of preferred shares, issuable in series, with no par value

Iss

wing table summarizes the changes in issued common shares of the Company:

For the years ended December 31,	**2013**		2012	
	Number of Shares	**Value**	Number of Shares	Value
Balance, beginning of year	**53,801,809**	**$391,120**	49,290,751	$339,178
Issued for cash (i)	-	-	2,875,000	32,877
Dividend reinvestment plan (ii)	**296,004**	**3,346**	347,473	4,367
Stock options exercised	**144,349**	**1,036**	471,593	4,195
Warrants exercised	-	-	816,992	10,000
Stock issued for acquisitions	**194,774**	**2,390**	-	-
Restricted share units redeemed	**21,838**	**289**	-	-
Tax benefit of issuance costs	-	-	-	503
Balance, end of year	**54,458,774**	**$398,181**	53,801,809	$391,120

(i) Net of issue costs of $Nil (2012 - $1,616)
(ii) Net of issue costs of $31 (2012 - $29)

Dividends

Killam paid monthly dividends as declared by the Board of Directors on or about the 15th day of each month. An annualized dividend of $0.58 was paid in 2013. The dividend increased to $0.60 on an annualized basis for the January 2014 dividend payable on February 17, 2014.

٢ الجريدة الرسمية - العدد ٢٢ (مكرر) فى ٢ يونية سنة ٢٠٠٢

قانون رقم ٨٢ لسنة ٢٠٠٢

بإصدار قانون حماية حقوق الملكية الفكرية

رئيس الجمهورية

قرر مجلس الشعب القانون الآتى نصه ، وقد أصدرناه :

(المادة الأولى)

يعمل بأحكام القانون المرافق فى شأن حماية حقوق الملكية الفكرية .

(المادة الثانية)

تلغى القوانين الآتية :

(أ) القانون رقم ٥٧ لسنة ١٩٣٩ بشأن العلامات والبيانات التجارية .

(ب) القانون رقم ١٣٢ لسنة ١٩٤٩ بشأن براءات الاختراع والرسوم والنماذج الصناعية ،
عدا أحكام براءات الاختراع الخاصة بالمنتجات الكيميائية المتعلقة بالأغذية
والمنتجات الكيميائية الصيدلية فتلغى اعتباراً من أول يناير سنة ٢٠٠٥

يصدر مجلس الوزراء بناء على عرض الوزراء المختصين اللائحة التنفيذية فى
مدة لا تزيد على شهر من تاريخ العمل بهذا القانون . ويصدر الوزراء المختصون كل فى حدود
اختصاصه القرارات اللازمة لتنفيذ أحكام القانون المرافق .

Fig. 8. Examples of the trained model's inferences on Arabic, Latin, and Chinese documents.

5 Conclusion and Prospects

We have introduced a new technique for analyzing document layouts, which involves utilizing the YOLOV7 model object detector. By training the model on the DocLayNet dataset, we could accurately detect and classify the various components of documents into one of eleven classes: caption, footnote, formula, list-item, page footer, page header, picture, section header, table, text, and title. This deep learning approach has proven to be efficient and effective, significantly improving the accuracy and efficiency of the base paper. The findings from this work indicate that object detection models can be utilized for document layout analysis, which presents a range of opportunities for automating document processing and analysis tasks. Future work could include further fine-tuning the model and incorporating additional image pre-processing techniques.

References

1. Bukhari, S.S., Breuel, T.M., Asi, A., El-Sana, J.: Layout analysis for Arabic historical document images using machine learning. In: 2012 International Conference on Frontiers in Handwriting Recognition, pp. 639–644. IEEE (2012)
2. Huang, Y., et al.: A yolo-based table detection method. In: 2019 International Conference on Document Analysis and Recognition (ICDAR), pp. 813–818. IEEE (2019)
3. Bethesda (MD): National Library of Medicine: PMC open access subset (2003). https://www.ncbi.nlm.nih.gov/pmc/tools/openftlist/. Accessed 15 Jan 2023
4. Minouei, M., Soheili, M.R., Stricker, D.: Document layout analysis with an enhanced object detector. In: 2021 5th International Conference on Pattern Recognition and Image Analysis (IPRIA), pp. 1–5. IEEE (2021)
5. O'Gorman, L.: The document spectrum for page layout analysis. IEEE Trans. Pattern Anal. Mach. Intell. **15**(11), 1162–1173 (1993)
6. Pfitzmann, B., Auer, C., Dolfi, M., Nassar, A.S., Staar, P.: Doclaynet: a large human-annotated dataset for document-layout segmentation. In: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 3743–3751 (2022)
7. Tran, H.T., Nguyen, N.Q., Tran, T.A., Mai, X.T., Nguyen, Q.T.: A deep learning-based system for document layout analysis. In: 2022 The 6th International Conference on Machine Learning and Soft Computing, pp. 20–25 (2022)
8. Wang, C.Y., Bochkovskiy, A., Liao, H.Y.M.: Yolov7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. arXiv preprint arXiv:2207.02696 (2022)
9. WongKinYiu: yolov7 (2022). https://github.com/WongKinYiu/yolov7/releases/download/v0.1/yolov7_training.pt. Accessed 18 Jan 2023
10. Zhang, D., Mao, R., Guo, R., Jiang, Y., Zhu, J.: Yolo-table: disclosure document table detection with involution. Int. J. Doc. Anal. Recognit. (IJDAR) **26**, 1–14 (2022)
11. Zhang, H., Xu, C., Shi, C., Bi, H., Li, Y., Mian, S.: HSCA-Net: a hybrid spatial-channel attention network in multiscale feature pyramid for document layout analysis. J. Artif. Intell. Technol. **3**(1), 10–17 (2023)
12. Zhong, X., Tang, J., Yepes, A.J.: Publaynet: largest dataset ever for document layout analysis. In: 2019 International Conference on Document Analysis and Recognition (ICDAR), pp. 1015–1022. IEEE (2019)
13. Zulfiqar, A., Ul-Hasan, A., Shafait, F.: Logical layout analysis using deep learning. In: 2019 Digital Image Computing: Techniques and Applications (DICTA), pp. 1–5. IEEE (2019)