

# ICDAR 2023 Competition on Robust Layout Segmentation in Corporate Documents

Christoph Auer<sup>[0000-0001-5761-0422]</sup>, Ahmed Nassar<sup>[0000-0002-9468-0822]</sup>,  
 Maksym Lysak<sup>[0000-0002-3723-6960]</sup>, Michele Dolfi<sup>[0000-0001-7216-8505]</sup>,  
 Nikolaos Livathinos<sup>[0000-0001-8513-3491]</sup>, and Peter Staar<sup>[0000-0002-8088-0823]</sup>

IBM Research  
 {cau,ahn,mly,dol,nli,taa}@zurich.ibm.com

**Abstract.** Transforming documents into machine-processable representations is a challenging task due to their complex structures and variability in formats. Recovering the layout structure and content from PDF files or scanned material has remained a key problem for decades. ICDAR has a long tradition in hosting competitions to benchmark the state-of-the-art and encourage the development of novel solutions to document layout understanding. In this report, we present the results of our *ICDAR 2023 Competition on Robust Layout Segmentation in Corporate Documents*, which posed the challenge to accurately segment the page layout in a broad range of document styles and domains, including corporate reports, technical literature and patents. To raise the bar over previous competitions, we engineered a hard competition dataset and proposed the recent DocLayNet dataset for training. We recorded 45 team registrations and received official submissions from 21 teams. In the presented solutions, we recognize interesting combinations of recent computer vision models, data augmentation strategies and ensemble methods to achieve remarkable accuracy in the task we posed. A clear trend towards adoption of vision-transformer based methods is evident. The results demonstrate substantial progress towards achieving robust and highly generalizing methods for document layout understanding.

**Keywords:** Document Layout Analysis · Machine Learning · Computer Vision · Object Detection · ICDAR Competition

## 1 Introduction

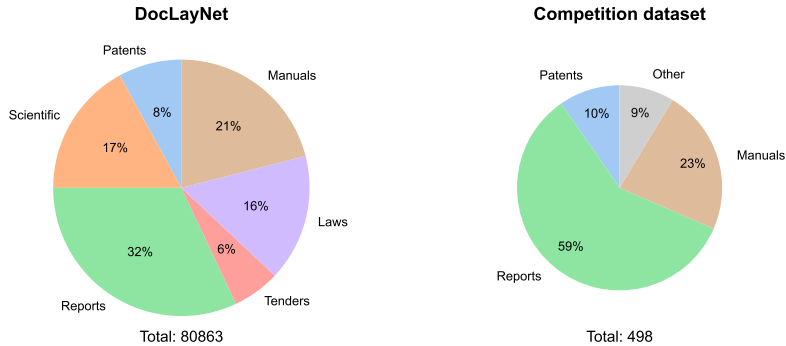
Document understanding is a key business process in the data-driven economy since documents are central to knowledge discovery and business insights. Converting documents into a machine-processable format is a particular challenge due to their huge variability in formats and complex structure. Recovering the layout structure and content from either PDF files or scanned material has remained a key problem since decades, and is as relevant-as-ever today. One can find vast amounts of approaches and solutions to this task [7,11,18,14,15,16], all of which are constrained to different degrees in the domains and document

styles that they can perform well on. A highly generalising model for structure and layout understanding has yet to be achieved.

ICDAR has organized various competitions in the past to benchmark the state-of-the-art and encourage the development of novel approaches and solutions to layout segmentation problems in documents [4,2,3,21]. In this report, we present the results of our *ICDAR 2023 Competition on Robust Layout Segmentation in Corporate Documents*, which posed the challenge to accurately segment the layout of a broad range of document styles and domains, including corporate reports, technical literature and patents. Participants were challenged to develop a method that could identify layout components in document pages as bounding boxes. These components include paragraphs, (sub)titles, tables, figures, lists, mathematical formulas, and several more. The performance of submissions was evaluated using the commonplace mean average precision metric (mAP) used in the COCO object detection competition [13]. To raise the bar over previous competitions, we proposed to use our recently published DocLayNet dataset [16] for model training, and engineered a challenging, multi-modal competition dataset with a unique distribution of new page samples.

Below, we present a detailed overview of this competition, including its datasets, evaluation metrics, participation, and results.

## 2 Datasets



**Fig. 1.** Dataset statistics of DocLayNet and the competition dataset.

### 2.1 Related Work

Layout segmentation datasets published in the recent past, such as PubLayNet [25] or DocBank [12], have enabled a big leap forward for ML-driven document understanding approaches due to their huge ground-truth size compared to earlier

work. However, these datasets still remain limited to a narrow domain of predominantly scientific documents, which is owed to their automatic ground-truth generation approach from mostly uniform XML or L<sup>A</sup>T<sub>E</sub>X sources. Despite exposing many different publisher layouts, all documents strongly share common traits and general structure. This has led to a saturation of ML model accuracy baselines at a very high level, with little room for improvement [21]. Yet, all publicly proposed ML models trained on these datasets generalize rather poorly to out-of-domain document samples, such as those found in the corporate world. For example, tables in invoices or manuals are difficult to detect correctly with models trained on scientific literature or books.

## 2.2 DocLayNet dataset

The DocLayNet dataset [16] addresses these known limitations by providing 80,863 page samples from a broad range of document styles and domains, which are fully layout annotated by human experts to a high-quality standard. DocLayNet is the first large-scale dataset covering a wide range of layout styles and domains, which includes Financial reports, Patents, Manuals, Laws, Tenders, and Technical Papers. It defines 11 class labels for rectangular bounding-box annotations, namely *Caption*, *Footnote*, *Formula*, *List-item*, *Page-footer*, *Page-header*, *Picture*, *Section-header*, *Table*, *Text* and *Title*. Detailed instructions and



**Fig. 2.** Select samples in the competition dataset (*Other* category) which fall outside of the layout distribution in DocLayNet.

guidance on how to consistently annotate the layout of DocLayNet pages were published in the accompanying layout annotation guideline.

Additionally, DocLayNet provides a JSON representation of each page with the original text tokens and coordinates from the programmatic PDF code. This opens the opportunity for new multi-modal ML approaches to the layout segmentation problem.

### 2.3 Competition dataset

To assess the layout segmentation performance of each team’s submissions, we engineered a competition dataset of 498 new pages in the same representation as the original DocLayNet dataset, which was provided to the participants without any annotation ground-truth. This competition dataset includes a mix of corporate document samples as shown in Fig. 2. Samples in the new *Other* category expose layouts which fall outside of the DocLayNet layout space.

## 3 Task

We designed the competition objective as a straightforward object detection task, since this is well-understood in the computer-vision community and fits the representation format of our DocLayNet dataset. Participants of our competition were challenged to develop methods that can identify layout components in document pages as rectangular bounding boxes, labelled with one of the 11 classes defined in the DocLayNet dataset (see Fig. 3). The performance of each team’s approach was evaluated on our competition dataset using the well established COCO mAP metric.

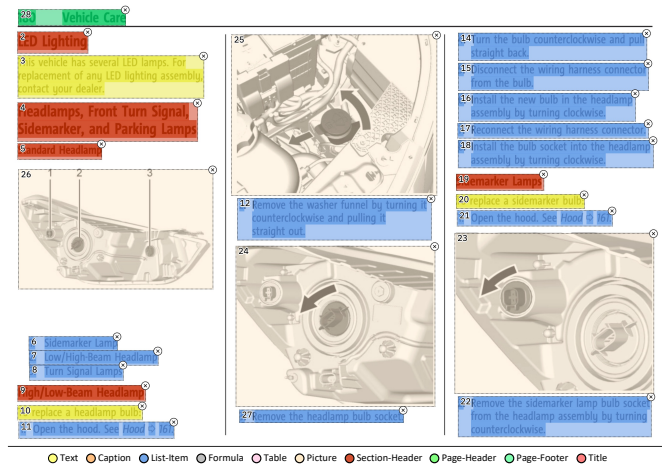


Fig. 3. Example page with bounding-box annotations.

**Submission format:** Since the COCO dataset format [13] and tooling is well established in the object detection community, we provided a standard COCO dataset file as part of our competition dataset, which includes the definition of class labels and image identifiers, but no ground-truth annotation data. Submissions were expected in the format of a JSON file complying with the commonly used *COCO results* schema, including complete bounding-box predictions for each page sample, matching to the identifiers defined in our provided dataset file.

**Evaluation metric:** All submissions were evaluated using the Mean Average Precision (mAP) @ Intersection-over-Union (IoU) [0.50:0.95] metric, as used in the COCO object detection competition. In detail, we calculate the average precision for a sequence of IoU thresholds ranging from 0.50 to 0.95 with a step size of 0.05, using the standard *pycocotools* library<sup>1</sup>. This metric was computed for every document category in the competition dataset separately. Then, the mean of the mAPs across all categories was computed with equal weights per category. The final ranking of every team’s submissions was based on the overall mAP.

## 4 Competition

**Schedule:** Our competition was officially announced on December 19th, 2022 and ended on April 3rd, 2023. The regular competition phase ended on March 26th, 2023 and the final week was run as a dedicated extension phase. Results of both phases are reflected in section 5.

**Setup:** We launched a competition website<sup>2</sup> to provide task descriptions, instructions, resources and news updates for the competition. For submission management, automatic online evaluation and tracking team submissions on a leader board, we relied on the free-to-use *EvalAI* platform [19]. To ensure fair conditions and prevent reverse engineering of our ground-truth, each team was originally granted 10 submission attempts on the evaluation platform. We increased this limit by 5 attempts for the extension phase. The feature in EvalAI to declare submissions private or public allowed teams to create multiple private submissions and check how they perform in evaluation before deciding to re-submit one of them as an official entry. The test-score for each submission was provided directly after submission. The latter has advantages and disadvantages. On the one hand, teams have a direct feedback on the quality of their results and can explore different strategies, which is one of the main motivations of this competition. On the other hand, it can also be used to overfit the model. For this explicit reason, we limited the number of submissions of each team to 10 (with extension 15). To set a baseline for the leader board, the competition organizers created an initial submission entry, which was visible to all teams.

<sup>1</sup> [pypi.org/project/pycocotools](https://pypi.org/project/pycocotools)

<sup>2</sup> [ds4sd.github.io/icdar23-doclaynet](https://ds4sd.github.io/icdar23-doclaynet)

## 5 Results

### 5.1 Overview

After the competition ended, we counted 45 team registrations, which altogether created 374 private or public submissions. Out of these, 21 team decided to make at least one public submission which counts towards the final ranking. Table 1 shows the results achieved by the participating teams for the regular submission phase and the extension phase of the competition. More detailed analysis and descriptions of selected methods from the participants are presented below.

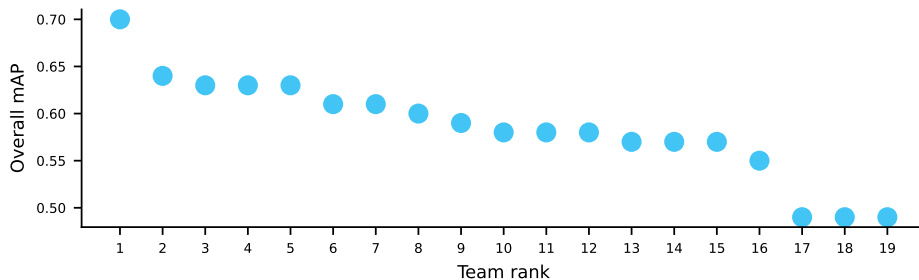
**Table 1.** Leaderboard of our competition with all teams ranking above our baseline (rank 19). Ranks are shown separately for the regular phase (*reg*) and the extension phase (*ext*)

Team	mAPs after extension					Ranking		
	Overall	Rep	Man	Pat	Other	reg.	ext.	Diff
docdog	<b>0.70</b>	<b>0.66</b>	<b>0.69</b>	0.84	<b>0.62</b>	1	1	↔
BOE_AIoT_CTO	0.64	0.54	0.67	0.84	0.52	2	2	↔
INNERCONV	0.63	0.57	0.63	<b>0.85</b>	0.48	3	3	↔
LC-OCR	0.63	0.61	0.65	0.77	0.48	5	4	↗1
DXM-DI-AI-CV-TEAM	0.63	0.54	0.63	0.82	0.51	4	5	↘1
alexsue	0.61	0.53	0.63	0.81	0.49	6	6	↔
PIX	0.61	0.52	0.63	0.82	0.46	-	7	*
Acodis	0.60	0.53	0.62	0.80	0.46	15	8	↗7
Linkus	0.59	0.49	0.64	0.77	0.48	7	9	↘2
TTW	0.58	0.49	0.61	0.80	0.42	8	10	↘2
amdoc	0.58	0.47	0.62	0.80	0.42	12	11	↗1
CVC-DAG	0.58	0.49	0.61	0.77	0.44	9	12	↘3
SPDB LAB	0.57	0.48	0.57	0.80	0.44	10	13	↘3
Alphastream.ai	0.57	0.47	0.57	0.79	0.45	11	14	↘3
Hisign	0.57	0.48	0.62	0.79	0.39	13	15	↘2
DLVC	0.55	0.53	0.57	0.74	0.38	14	16	↘2
Vamshikancharla	0.49	0.36	0.48	0.76	0.37	16	17	↘1
Azure	0.49	0.44	0.55	0.59	0.38	17	18	↘1
ICDAR23 DocLayNet organizers (Baseline)	0.49	0.38	0.52	0.70	0.35	18	19	↘1

### 5.2 General analysis

**Layout segmentation performance:** It is apparent that the top-ranking team (docdog) has presented a solution that is performing notably superior compared to the remainder of the field, as evidenced by their 6% lead in total mAP score over the second best submission. This result is achieved through outperforming every other team in the *Reports* category (5% lead) and the particularly difficult *Others* category (10% lead). From the second rank down, we observe a very

competitive field with many teams achieving similar levels of mAP performance, ranging from 0.64 (rank 2) to 0.55 (rank 16). Two more teams ranked just slightly above our baseline mAP of 0.49 (see Fig. 4).



**Fig. 4.** Distribution of overall mAP achieved by teams. Numbers and ranks refer to extension phase.

Throughout the extension phase, we observed mostly small improvements of overall mAP within a 1-2% range, with few exceptions such as team *docdog* and team *Acodis*, which managed to improve by 4% and 6% over their result from the regular competition phase, respectively. Team *PIX* joined as a new entrant in the extension phase only.

The highest, and also most consistent performance across submissions, is observed in the *Patent* category, with 12 teams achieving an mAP of 0.79 or better. This is consistent with our expectations, since Patent document layouts are the most uniform and structured. The diverse, free-style layouts in the *Reports* and *Others* categories posed a considerably bigger challenge, with mAPs generally ranging in the low 60%s and 50%s respectively.

Two interesting observations can be made. On one hand, we find significantly lower mAPs in the submissions across our competition set categories than those which were achieved for example on PubMed Central papers in the ICDAR 2021 Competition on Scientific Literature Parsing[21]. This can be attributed both to the more challenging layouts and higher class count of the DocLayNet dataset, as well as to the distribution bias and hard samples we engineered the competition dataset to expose. On the other hand, we see a significant spread of mAPs across the final submissions, with almost all teams exceeding the baseline by a significant margin. This delivers evidence that the participating teams have created solutions that differentiate themselves significantly from previous off-the-shelf object detection methods (see baseline). It also shows that the investment to develop sophisticated methods is beneficial to obtain superior performance on this dataset.

**Models and Strategies:** In the solutions presented by the top five teams, we were pleased to see novel and interesting combinations of recent computer

vision models, data augmentation strategies and ensemble methods applied to solve the layout segmentation task with high accuracy. All top-ranking solutions adopt, to different degrees, the recently emerging deep-learning models based on vision transformer methods and self-supervised pre-training, such as the generic DINO[22] and MaskDINO[10] models, or the document-understanding focused DiT[11] and LayoutLMv3[7] models. The DocLayNet dataset was used for fine-tuning in this context. Several solutions combine these new-generation vision models with more traditional, CNN-based object detectors such as YOLO [8] through model ensemble, for example through *Weighted Boxes Fusion* [17]. Data augmentation strategies used by the teams include multi-scale and mosaic methods, as well as deriving synthetic datasets from DocLayNet. Two of the top five teams reported that they include the additional text cell layer provided by DocLayNet and the competition dataset in their approach. No teams stated to create private ground-truth data that was not derived from DocLayNet.

### 5.3 Method descriptions

Below we summarize the methods reported by the top five teams for comparison reasons to our best understanding. We would like to extend our thanks to all competition teams who took the time to provide us with a comprehensive description of their methods.

#### Team doccdog (Tencent WeChat AI)

The team created a synthetic image dataset of 300,000 samples based on the training dataset. For the task of layout prediction, the team used two models, YOLOv8 [8] and DINO [22]. An extra classification model was trained to categorize the samples of the competition dataset into the document categories. YOLOv8 models with different network sizes (medium, large, x-large) were trained, each with different input resolutions, for ensemble and optimization of the detection performance. For the DINO model, the team applied a carefully designed augmentation strategy and integrated focal modulation networks [20] in the backbone for improved performance. Separate models were trained per category, both with and without synthetic data. Model hyper-parameters were optimized using a Tree-Structured Parzen Estimator (TPE) [1] to find the best weights. Prediction results from the individual models were combined using Weighted Boxes Fusion (WBF) [17] and fine-tuned using text cell coordinates from the JSON representation of the samples in the competition dataset. Further detail on the approach is provided in the team’s *WeLayout* paper [23].

#### Team BOE\_AIoT\_CTO

The team relied exclusively on the DocLayNet dataset for training, and applied scale and mosaic methods for image augmentation. For the task of layout prediction, the team trained two object detection models, YOLOv5 [9] and



YOLOv8 [8]. Training was conducted over 150 epochs using BCELoss with Focal-Loss, and mosaic augmentation was cancelled for the final 20 epochs. Additionally, a DiT model [11] (dit-large) was fine-tuned using the DocLayNet dataset. To improve vertical text detection, the team added multi-scale image training. Predictions for the final submission were ensembled from three detectors to achieve superior performance.

### Team INNERCONV

For the task of layout prediction, the team uses the MaskDINO model [10]. MaskDINO is derivative of DINO [22] which introduces a mask prediction branch in parallel to the box prediction branch of DINO. It achieves better alignment of features between detection and segmentation. In training, only the image representation of the DocLayNet dataset is used. In inference, the team applied the Weighted Boxes Fusion (WBF) technique [17] to ensemble the predictions on multiple scales of the same input image.

### Team LC-OCR (CVTE)

For the task of layout prediction, the team applied two models, VSR [24] and LayoutLMv3 [7], which use pre-trained weights. Prediction results from both models are merged in inference. Detections for the classes *Footnote*, *Picture*, *Table* and *Title* were taken from LayoutLMv3, the remainder of classes from VSR. In VSR, the team included the text cell information provided in the JSON representation of DocLayNet.

### Team DXM-DI-AI-CV-TEAM (Du Xiaoman Financial)

For the task of layout prediction, the team trained different versions of Cascade Mask R-CNN [6] models, based on a DiT[11] backbone (DiT-large), and fuse prediction results using different models.

### Baseline of ICDAR 2023 DocLayNet organizers

To set a comparison baseline for the competition, the organizers used a YOLOv5 model (medium size), and trained it solely on the DocLayNet training dataset, with images re-scaled to square 1024 by 1024 pixels. The model was trained from scratch with default settings for 80 epochs. We applied standard augmentation techniques such as mosaic, scale, flipping, rotation, mix-up and image levels.

## 6 Conclusions

We believe that this ICDAR competition served its purpose well to benchmark the state-of-the-art solutions to the layout segmentation task in documents, and

again encouraged the development of unique new approaches. Our new competition dataset was designed to raise the bar over previous competitions by providing diverse, challenging page layouts, paired with multi-modal representation. This enabled participants to test the generalization power of the latest computer-vision methods, especially with recently emerging models based on self-supervised pre-training and visual transformers.

We were pleasantly surprised by the high level of engagement in this competition, with 45 teams registering, out of which 21 teams created an official final submission. The budget of 15 total submissions was fully used by the majority of the contestants. Overall, the level of sophistication demonstrated in the approaches went well beyond our anticipation. One core take-away is the importance of data augmentation and ensemble techniques to improve the layout prediction performance beyond the level of what any single end-to-end model currently delivers. It was also interesting to observe how the various techniques applied by the different teams in many cases yielded similar results in overall accuracy. The remarkable progress demonstrated by the top-performing teams in this competition will be valuable for future research on highly capable document understanding models.

We are also glad to see this competition spark wider interest in the community, as it prompted some members to build and share fully runnable example codes and publish blog articles on training and inference with DocLayNet and pre-trained models[5]. To support these community efforts, we made DocLayNet available on the HuggingFace datasets hub<sup>3</sup>. As such, we believe that this ICDAR competition has also helped to establish the DocLayNet dataset as a well known asset for document understanding research and applications.

## Acknowledgements

We would like to thank all participants for their remarkable efforts and contributions to this competition, and the Competitions Chairs for providing the opportunity to host this competition in ICDAR 2023.

## References

1. Bergstra, J., Bardenet, R., Bengio, Y., Kégl, B.: Algorithms for hyper-parameter optimization. In: Shawe-Taylor, J., Zemel, R., Bartlett, P., Pereira, F., Weinberger, K. (eds.) *Advances in Neural Information Processing Systems*. vol. 24. Curran Associates, Inc. (2011)
2. Clausner, C., Antonacopoulos, A., Plotschacher, S.: Icdar2017 competition on recognition of documents with complex layouts - rdcl2017. In: *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*. vol. 01, pp. 1404–1410 (2017). <https://doi.org/10.1109/ICDAR.2017.229>
3. Déjean, H., Meunier, J.L., Gao, L., Huang, Y., Fang, Y., Kleber, F., Lang, E.M.: ICDAR 2019 Competition on Table Detection and Recognition (cTDaR) (Apr 2019). <https://doi.org/10.5281/zenodo.2649217>

<sup>3</sup> <https://huggingface.co/datasets/ds4sd/DocLayNet>

4. Göbel, M., Hassan, T., Oro, E., Orsi, G.: Icdar 2013 table competition. In: 2013 12th International Conference on Document Analysis and Recognition. pp. 1449–1453 (2013). <https://doi.org/10.1109/ICDAR.2013.292>
5. Guillou, P.: Document ai | processing of doclaynet dataset to be used by layout models of the hugging face hub (finetuning, inference) (Jan 2023), [https://medium.com/@pierre\\_guillou/308d8bd81cdb](https://medium.com/@pierre_guillou/308d8bd81cdb)
6. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask r-cnn (2018)
7. Huang, Y., Lv, T., Cui, L., Lu, Y., Wei, F.: Layoutlmv3: Pre-training for document ai with unified text and image masking (2022)
8. Jocher, G., Chaurasia, A., Qiu, J.: YOLO by Ultralytics (Jan 2023), <https://github.com/ultralytics/ultralytics>
9. Jocher, G., Chaurasia, A., Stoken, A., Borovec, J., NanoCode012, Kwon, Y., Michael, K., TaoXie, Fang, J., imyhxy, Lorna, Yifu, Z., Wong, C., V, A., Montes, D., Wang, Z., Fati, C., Nadar, J., Laughing, UnglvKitDe, Sonck, V., tkianai, yxNONG, Skalski, P., Hogan, A., Nair, D., Strobel, M., Jain, M.: ultralytics/yolov5: v7.0 - YOLOv5 SOTA Realtime Instance Segmentation (Nov 2022). <https://doi.org/10.5281/zenodo.7347926>
10. Li, F., Zhang, H., xu, H., Liu, S., Zhang, L., Ni, L.M., Shum, H.Y.: Mask dino: Towards a unified transformer-based framework for object detection and segmentation (2022)
11. Li, J., Xu, Y., Lv, T., Cui, L., Zhang, C., Wei, F.: Dit: Self-supervised pre-training for document image transformer. In: ACM Multimedia 2022 (October 2022)
12. Li, M., Xu, Y., Cui, L., Huang, S., Wei, F., Li, Z., Zhou, M.: Docbank: A benchmark dataset for document layout analysis. In: Proceedings of the 28th International Conference on Computational Linguistics. pp. 949–960. COLING, International Committee on Computational Linguistics (dec 2020). <https://doi.org/10.18653/v1/2020.coling-main.82>, <https://aclanthology.org/2020.coling-main.82>
13. Lin, T.Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C.L., Dollár, P.: Microsoft coco: Common objects in context (2015)
14. Livathinos, N., Berrospi, C., Lysak, M., Kuropiatnyk, V., Nassar, A., Carvalho, A., Dolfi, M., Auer, C., Dinkla, K., Staar, P.: Robust pdf document conversion using recurrent neural networks. Proceedings of the AAAI Conference on Artificial Intelligence **35**(17), 15137–15145 (May 2021), <https://ojs.aaai.org/index.php/AAAI/article/view/17777>
15. Nassar, A., Livathinos, N., Lysak, M., Staar, P.: Tableformer: Table structure understanding with transformers. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 4614–4623 (June 2022)
16. Pfitzmann, B., Auer, C., Dolfi, M., Nassar, A.S., Staar, P.W.J.: Doclaynet: A large human-annotated dataset for document-layout segmentation. In: Zhang, A., Rangwala, H. (eds.) KDD '22: The 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, August 14 - 18, 2022. pp. 3743–3751. ACM (2022). <https://doi.org/10.1145/3534678.3539043>
17. Solovyev, R., Wang, W., Gabruseva, T.: Weighted boxes fusion: Ensembling boxes from different object detection models. Image and Vision Computing **107**, 104117 (2021). <https://doi.org/10.1016/j.imavis.2021.104117>
18. Staar, P.W.J., Dolfi, M., Auer, C., Bekas, C.: Corpus conversion service: A machine learning platform to ingest documents at scale. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. p.

- 774–782. KDD '18, Association for Computing Machinery, New York, NY, USA (2018). <https://doi.org/10.1145/3219819.3219834>
19. Yadav, D., Jain, R., Agrawal, H., Chattopadhyay, P., Singh, T., Jain, A., Singh, S.B., Lee, S., Batra, D.: Evalai: Towards better evaluation systems for ai agents (2019)
  20. Yang, J., Li, C., Dai, X., Gao, J.: Focal modulation networks. In: Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., Cho, K., Oh, A. (eds.) *Advances in Neural Information Processing Systems*. vol. 35, pp. 4203–4217. Curran Associates, Inc. (2022)
  21. Yepes, A.J., Zhong, P., Burdick, D.: Competition on scientific literature parsing. In: *Proceedings of the International Conference on Document Analysis and Recognition*. pp. 605–617. ICDAR, LNCS 12824, Springer-Verlag (sep 2021). [https://doi.org/10.1007/978-3-030-86337-1\\_40](https://doi.org/10.1007/978-3-030-86337-1_40)
  22. Zhang, H., Li, F., Liu, S., Zhang, L., Su, H., Zhu, J., Ni, L.M., Shum, H.Y.: Dino: Detr with improved denoising anchor boxes for end-to-end object detection (2022)
  23. Zhang, M., Cao, Z., Liu, J., Niu, L., Meng, F., Zhou, J.: Welayout: Wechat layout analysis system for the icdar 2023 competition on robust layout segmentation in corporate documents (2023), <https://arxiv.org/abs/2305.06553>
  24. Zhang, P., Li, C., Qiao, L., Cheng, Z., Pu, S., Niu, Y., Wu, F.: Vsr: A unified framework for document layout analysis combining vision, semantics and relations (2021)
  25. 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 (Sep 2019). <https://doi.org/10.1109/ICDAR.2019.00166>