

Supplementary Material – UniTabBank: A Large Scale Multi-Lingual, Multi-Layout, Multi-Type, Multi-Format Dataset for Table Detection

Ajoy Mondal Saumya Mundra Avijit Dasgupta C. V. Jawahar
CVIT, IIIT, Hyderabad, India

{ajoy.mondal, jawahar}@iiit.ac.in, {saumya.mundra, avijit.dasgupta}@research.iiit.ac.in

In this supplementary material, we discuss the following

- List of document sources (Appendix A).
- Language-wise statistics and additional visual examples of the **UniTabBank** dataset (Appendix B).
- Detailed discussion of evaluation metrics (Appendix C).
- Additional experiments and analysis on the generalization capability of **UniTabBank** (Appendix D).
- Additional quantitative results (Appendix E).
- Additional visual results (Appendix F).
- Discussion of further experiments (Appendix G).

A. List of Sources

- https://www.arig.net/2017/?post_type=annualreports
- https://www.arabpotash.com/En/List/Annual_Reports
- https://www.dvc.gov.in/cms-web/annual_report
- https://www.cug.ac.in/annual_report.php
- <https://dopt.gov.in/hi/reports/annual-report>
- <https://punjabandsindbank.co.in/content/archive-annual>
- <https://rural.gov.in/publications/annual-report>
- <https://ypsa.org/annual-reports/>
- <https://www.bb.org.bd/en/index.php/publication/publicctn/0/2>
- <https://gshrc.gujarat.gov.in/annual-reports.htm>
- <https://samagrashiksha.ssagujarat.org/en/Annual-Report>
- <https://www.sumul.com/sumul-annual-report.html>
- <https://www.kamdhenuni.edu.in/annual-reports>
- <https://samagrashiksha.ssagujarat.org/en/Annual-Report>
- <https://gswtindia.com/gswt-annual-reports/>
- <https://kerc.karnataka.gov.in/37/annual-report/kn>
- <https://www.keralapsc.gov.in/ml/annual-report>
- <https://startupmission.kerala.gov.in/reports/annual>
- <https://www.dnyandeep.co.in/marathi/annual-report.html>
- <https://www.shivdaulatbank.com/annual-report>
- <https://dbatu.ac.in/annual-reports/>
- <https://pc.odisha.gov.in/publication/activity-reports>
- <https://dhsodisha.nic.in/?q=node/103>
- <https://industries.odisha.gov.in/publication/annual-reports>
- <https://www.kyoto-u.ac.jp/ja/about/public/issue/annual-report>
- <https://www.jica.go.jp/english/about/disc/report/index.html>
- <https://www.cnoocld.com/tzzgx/dqbg/nb/>
- <https://www.aerothai.co.th/th/about/document/annual-report>
- <https://www.businesseventsthailand.com/th/about-us/annual-reports>
- <https://www.dusit-international.com/en/investor-relations/document/annual-reports>
- <https://www.unicef.org/fr/topics/rapport-annuel>
- <https://vietnam.actionaid.org/en/annual-reports>
- <https://www.economica.vn/docs/publications-and-reports/17>
- <https://www.ina.go.id/annual-report>
- <https://indorama.co.id/in/bagian-investor/laporan-tahunan.html>

Language	Image				Language	Table			
	Total number	Training number	Validation number	Test number		Total number	Training number	Validation number	Test number
Arabic	1413	847	143	422	Arabic	2328	1371	231	775
Assamese	390	233	41	115	Assamese	481	295	48	142
Bengali	2673	1602	271	797	Bengali	3581	2180	359	1040
Bodo	144	86	15	43	Bodo	184	111	18	53
Chinese	1514	908	152	451	Chinese	2318	1367	265	732
English	3824	2292	387	1137	English	5803	3484	554	1791
Farsi	12	7	2	3	Farsi	18	8	2	8
French	3429	2057	344	1027	French	5496	3315	548	1662
Gujarati	4857	2912	490	1440	Gujarati	6927	4175	690	2097
Hindi	10300	6179	1033	3151	Hindi	16830	10068	1733	5468
Indonesia	1572	943	158	469	Indonesia	2024	1223	215	608
Japanese	1179	706	121	350	Japanese	1955	1186	197	583
Kannada	5402	3240	543	1623	Kannada	8777	5339	821	2690
Korean	178	106	20	52	Korean	369	221	44	106
Malayalam	1471	881	151	439	Malayalam	1865	1114	190	603
Manipuri	74	44	8	22	Manipuri	90	53	10	29
Marathi	2615	1567	266	775	Marathi	3778	2257	388	1159
Nepali	683	409	70	200	Nepali	901	525	94	283
Oriya	1758	1051	183	524	Oriya	2674	1571	262	886
Punjabi	2703	1620	273	806	Punjabi	3304	1993	325	995
Sanskrit	96	57	11	28	Sanskrit	107	66	12	30
Sinhala	793	475	81	237	Sinhala	993	599	113	285
Spanish	400	240	40	119	Spanish	501	314	45	145
Tamil	2676	1603	273	794	Tamil	3434	2042	370	1037
Telugu	1310	785	134	387	Telugu	1731	1068	169	510
Thai	1417	850	142	424	Thai	1806	1053	187	601
Urdu	1065	637	110	314	Urdu	1322	793	134	403
Vietnamese	1495	896	151	448	Vietnamese	1582	947	157	478
Total	55443	33233	5613	16597	Total	82114	48738	8181	25195

Table 1. Language wise statistics of UniTabBank dataset.

- <https://solusibangunindonesia.com/laporan-tahunan/>
- <https://www.ksei.co.id/annual-reports?setLocale=en-US>
- <https://www.antofagasta.co.uk/investors/annual-report-library/>
- <https://www.octapharma.com/about-us/who-we-are/annual-report>

B. Language-wise Statistics and Visual Examples

Language-wise Table Frequencies of UniTabBank dataset: Table 1 summarizes the overall table frequencies with respect to language and their distribution in each set.

More Visual Samples from UniTabDet Dataset: Fig. 1 shows examples with diverse document types, formats, layouts, and languages, from our UniTabBank dataset.

C. Evaluation Metrics

The Precision (P), Recall (R), and F1 score (F1) [2, 6, 21] are computed as

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{F1 score} = \frac{2 \times P \times R}{P + R},$$

where TP , TN , FP , FN refer to true positive, true negative, false positive, and false negative, respectively.

Weighted average F1 score [6, 22] over varying IoU is calculated as

$$\text{Weighted Average } F1 = \frac{\sum_{i=1}^t IoU_i \cdot F1 @ IoU_i}{\sum_{i=1}^t IoU_i}$$

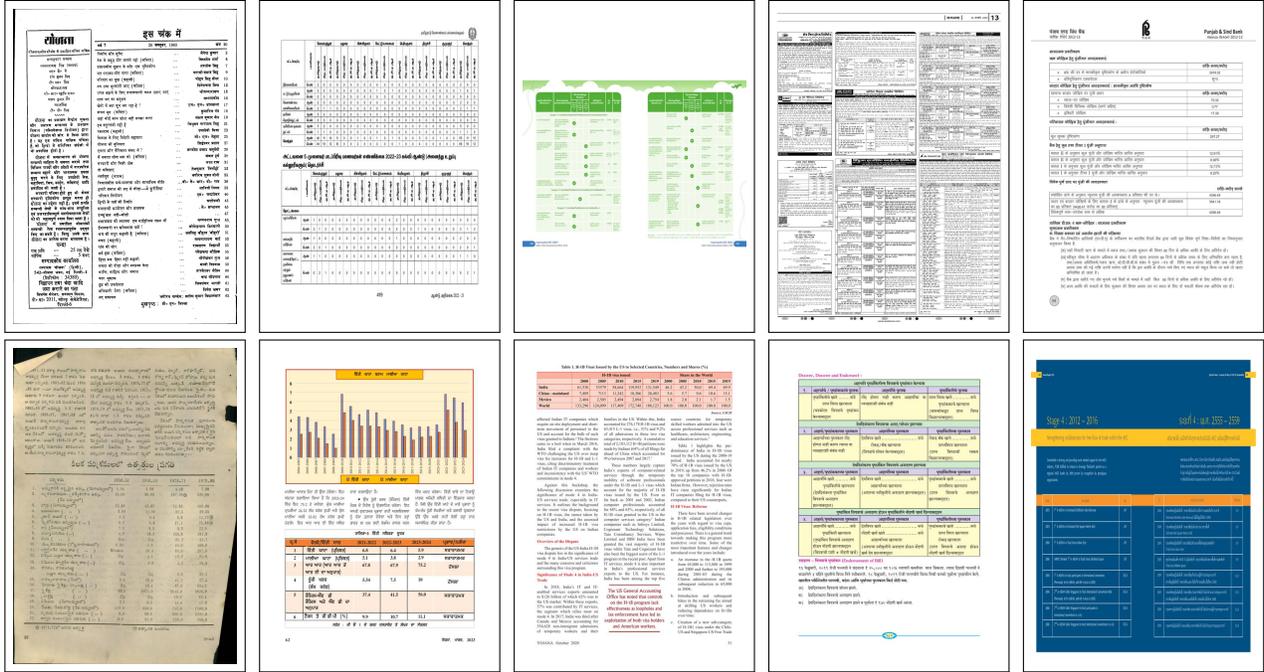


Figure 1. Shows examples of images from our UniTabBank dataset with diverse document types, formats, layouts, and languages.

We use standard average precision (AP) metrics from the object detection literature [8, 10, 11, 16], including AP_{50} and AP_{75} , which correspond to average precision at IoU thresholds of 0.50 and 0.75, respectively, as well as the mean AP computed over the IoU range [0.50–0.95]. These metrics provide a comprehensive assessment of table detection [7, 20] performance across varying localization strictness.

D. Generalization Capability of UniTabBank

The cross-benchmark evaluation reported in Table 2 highlights a clear trade-off between in-domain accuracy and cross-domain generalization. Models trained and tested on the same dataset achieve the best results (e.g., $ICT-TD \rightarrow ICT-TD$: $AP=0.949$, $TNCR \rightarrow TNCR$: $AP=0.956$, $ICDAR2019 \rightarrow ICDAR2019$: $AP=0.835$), underscoring strong dataset-specific learning. However, training on PubTables or TableBank leads to poor transferability, with sharp performance drops on unseen datasets (e.g., $PubTables \rightarrow ICDAR2019$: $AP=0.371$, $TableBank \rightarrow Marmot$: $AP=0.492$). In contrast, UniTabBank-trained UniTabDet consistently achieves strong and stable performance across diverse benchmarks ($AP=0.849$ on $ICT-TD$, $AP=0.913$ on $TNCR$, $AP=0.750$ on $ICDAR2019$, $AP=0.644$ on $Marmot$, and $AP=0.912$ on $ICDAR2013$), often ranking second only to in-domain training. It demonstrates UniTabBank’s strength in capturing document format, layout, type, and linguistic diversity, enabling broader generalization than ex-

isting benchmarks.

E. More Quantitative Results

Comparison with SOTA on TD Benchmarks Table 3, our fine-tuned UniTabDet[†] outperforms prior approaches by nearly 9% at IoU 0.95 on the $ICT-TD$ dataset, demonstrating more precise table localization. Table 4 highlights a 2% F1 improvement over the state-of-the-art SparseTableDet [22] on the $TNCR$ dataset using UniTabDet[†], while even the non-finetuned version performs comparably to leading methods. In Table 5 shows that UniTabDet outperforms the state-of-the-art method by 18.4% on the $STDW$ dataset. Table 6 summarizes the performance comparison on the $Marmot$ dataset, where UniTabDet achieves a 1.5% higher F1 score than state-of-the-art methods. For the $ICDAR-2013$ dataset, as shown in Table 7, most existing models, including UniTabDet, achieve 100% accuracy. Similarly, Table 8 presents the results on the $ICDAR-2019$ dataset, where UniTabDet, without fine-tuning on $ICDAR-2019$, performs 18% lower in F1 score compared to the state-of-the-art SparseTableDet. This drop is primarily due to the presence of handwritten tables in the dataset. Table 9 shows that UniTabDet achieves a 9% higher F1 score than existing methods on the $UNLV$ dataset at IoU 0.5. These results highlight that UniTabDet, trained on UniTabBank, demonstrates strong generalization and performs robustly across diverse table detection benchmarks, regardless of document type, format, language, or layout.

Training Set	Test Set	AP ₅₀	AP ₇₅	AP
PubTables	ICT-TD	0.711	0.558	0.497
TableBank		0.764	0.726	0.704
UniTabBank		0.903	0.858	0.849
ICT-TD		0.984	<u>0.963</u>	<u>0.949</u>
TNCR		0.892	0.858	0.841
ICDAR2019		<u>0.983</u>	0.969	0.950
PubTables	TNCR	0.717	0.583	0.503
TableBank		0.866	0.845	0.764
UniTabBank		<u>0.971</u>	<u>0.954</u>	<u>0.913</u>
ICT-TD		0.963	0.946	0.894
TNCR		0.995	0.995	0.956
ICDAR2019		0.969	0.947	0.895
PubTables	ICDAR-2019	0.486	0.424	0.371
TableBank		0.536	0.447	0.447
UniTabBank		<u>0.865</u>	<u>0.790</u>	<u>0.750</u>
ICT-TD		0.658	0.505	0.498
TNCR		0.784	0.645	0.622
ICDAR2019		0.873	0.855	0.835
PubTables	Marmot	0.674	0.292	0.329
TableBank		0.790	0.510	0.492
UniTabBank		0.920	0.722	0.644
ICT-TD		0.901	0.699	0.628
TNCR		<u>0.905</u>	<u>0.706</u>	<u>0.631</u>
ICDAR2019		0.904	0.697	0.630
PubTables	ICDAR-2013	0.930	0.929	<u>0.906</u>
TableBank		0.942	0.859	0.740
UniTabBank		<u>0.982</u>	0.956	0.912
ICT-TD		0.977	0.948	0.835
TNCR		0.972	0.923	0.794
ICDAR2019		0.983	<u>0.954</u>	0.842

Table 2. Cross-benchmark evaluation of UniTabDet trained on different datasets and tested across multiple benchmarks. Models trained on benchmark-specific datasets achieve high in-domain accuracy but generalize poorly, whereas UniTabBank-trained models achieve consistently strong cross-domain performance. Bold and underlined values represent the best and second-best results, respectively.

Model	Test: <i>ICT-TD</i>				
	F1(0.80)	F1(0.85)	F1(0.90)	F1(0.95)	WF1
TableDet [23]	0.936	0.916	0.891	0.757	0.871
DiffusionDet [23]	0.955	0.942	0.911	0.764	0.889
Def-DETR [23]	0.950	0.939	0.912	0.830	0.905
SparseR-CNN [23]	0.943	0.930	0.904	0.788	0.888
SparseTableDet [22]	0.972	0.964	0.942	<u>0.818</u>	<u>0.921</u>
UniTabDet	0.861	0.859	0.855	0.826	0.874
UniTabDet [†]	<u>0.961</u>	<u>0.960</u>	0.957	0.929	0.951

Table 3. Performance comparison on the *ICT-TD* dataset using F1 scores at multiple IoU thresholds (0.80, 0.85, 0.90, and 0.95), along with the weighted average F1 score. † indicates models fine-tuned on the *ICT-TD* dataset. Bold and underlined values denote the best and second-best results, respectively.

Model	Train		Test: <i>TNCR</i>		
	Dataset	#Image	P	R	F1
C Mask RCNN [1]	TNCR	6600	0.780	0.898	0.844
C RCNN [1]	TNCR	6600	0.798	0.891	0.841
Def-DETR [1]	TNCR	6600	0.867	0.896	0.881
SparseTableDet [22]	TNCR	6600	0.906	0.983	<u>0.943</u>
UniTabDet	UniTabBank	55,443	<u>0.934</u>	0.921	0.927
UniTabDet [†]	TNCR	6600	0.964	<u>0.970</u>	0.967

Table 4. Performance comparison on the *TNCR* dataset in terms of precision (P), recall (R), and F1 score at IoU thresholds ranging from 0.50 to 0.95. † indicates models fine-tuned on the *TNCR* dataset. C indicates Cascade. Bold and underlined values indicate the best and second-best results, respectively.

Model	Train		Test: <i>STDW</i>
	Dataset	#Image	AP ₅₀
RetinaNet [7]	-	-	<u>0.780</u>
UniTabDet	UniTabBank	55,443	0.964

Table 5. Performance comparison on the *STDW* dataset using average precision (AP) at IoU=0.5. Bold and underlined values represent the best and second-best results, respectively.

Model	Test: <i>Marmot</i>		
	P	R	F1
DeepDeSRT [17]	0.959	0.874	0.914
DeCNT [19]	0.850	0.946	0.895
CascadeTabNet [15]	0.911	0.915	0.905
CDeC-Net [2]	<u>0.975</u>	0.930	0.952
TableNet [13]	0.955	0.950	0.955
DETR [20]	0.953	0.954	0.914
TransTab [21]	0.965	<u>0.959</u>	<u>0.961</u>
UniTabDet	0.987	0.985	0.986

Table 6. Performance evaluation on the *Marmot* dataset in terms of precision (P), recall (R), and F1 score at IoU = 0.5. Bold and underlined values indicate the best and second-best results, respectively.

Performance Analysis based on Document-Type: Table 10 presents the performance comparison across different document types. Among them, textbooks typically have the simplest layouts, resulting in the highest average precision (AP) for all models. In contrast, newspapers feature more complex and varied layouts, leading to the lowest AP scores, highlighting the increased difficulty of table detection in such formats.

Performance Analysis based on Language: Table 11 presents the performance comparison across different languages. Since existing model Table-Transformer [20] trained with PubTables-1M — research articles in english is unable to obtain similar accuracy across languages. document in particular language available in all document types.

Model	Test: ICADR-2013		
	P	R	F1
CascadeTabNet [15]	1.000	1.000	1.000
TableDet [5]	1.000	1.000	1.000
DeCNT [19]	<u>0.996</u>	<u>0.996</u>	<u>0.996</u>
DeepDeSRT [17]	0.962	0.974	0.968
TableNet [13]	0.970	0.963	0.966
YOLOv3-TD [9]	0.949	1.000	0.973
SparseTableDet [22]	1.000	1.000	1.000
CDeC-Net [2]	1.000	1.000	1.000
TransTab [21]	0.981	0.978	0.962
UniTabDet	1.000	1.000	1.000

Table 7. Performance comparison on the ICADR-2013 dataset using precision (P), recall (R), and F1 score at IoU=0.5. Bold and underlined values represent the best and second-best results, respectively.

Model	Test: ICADR-2019				
	F1(60)	F1(70)	F1(80)	F1(90)	WF1
TableDet [5]	0.969	<u>0.969</u>	<u>0.957</u>	<u>0.919</u>	<u>0.951</u>
TableRadar [6]	0.969	0.957	0.951	0.897	0.940
NLPR-PAL [6]	<u>0.979</u>	0.966	0.939	0.850	0.927
CascadeTabNet [15]	0.943	0.934	0.925	0.901	0.901
SparseTableDet [22]	0.993	0.991	0.989	0.963	0.983
UniTabDet	0.829	0.816	0.799	0.785	0.805

Table 8. Performance evaluation on ICADR-2019 using F1 scores at multiple IoU thresholds (0.60, 0.70, 0.80, and 0.90), along with the weighted average F1 score.

Model	Train		Test: UNLV		
	Dataset	#Image	P	R	F1
DeCNT [19]	D4	4,622	0.786	<u>0.749</u>	0.767
CDeC-Net [2]	D4	4,622	<u>0.852</u>	0.736	<u>0.794</u>
UniTabDet	UniTabBank	55,443	0.911	0.851	0.880

Table 9. Performance comparison on the UNLV dataset in terms of precision (P), recall (R), and F1 score at IoU=0.5. D4 refers to the combined curation of ICADR-2013, ICADR-2017, and Mar-mot datasets. Bold and underlined values represent the best and second-best results, respectively.

Model	AP: Document-Type			
	AR	Book	Magazine	Newspaper
Table-Transformer [20]	0.808	0.809	0.753	0.127
SparseTableDet [22]	0.922	0.931	0.836	0.495
UniTabDet	0.956	0.985	0.964	0.912

Table 10. Performance comparison in terms of Average Precision (AP) at IoU [0.5–0.95] across different document types — Annual Reports (AR), Books, Magazines, and Newspapers in UniTabBank dataset.

While UniTabDet trained with UniTabBank — mixture of all languages and document types obtain similar AP across

Language	Table-Transformer [20]	SparseTableDet [22]	UniTabDet
Arabic	0.4614	0.805	0.978
Assamese	0.5578	0.917	0.954
Bengali	0.4868	0.781	0.912
Bodo	0.5573	0.944	0.994
Chinese	0.5070	0.851	0.975
English	0.5024	0.834	0.981
Farsi	0.2668	1.000	<u>0.995</u>
French	0.6114	0.855	0.903
Gujarati	0.5546	0.920	0.970
Hindi	0.5211	0.889	0.966
Indonesia	0.6341	0.816	0.931
Japanese	0.5443	0.910	0.959
Kannada	0.4801	0.924	0.942
Korean	0.6837	0.953	0.992
Malayalam	0.4462	0.889	0.959
Manipuri	0.4873	0.967	<u>0.995</u>
Marathi	0.5021	0.895	0.976
Nepali	0.4317	0.943	0.989
Oriya	0.3123	0.750	0.960
Punjabi	0.4824	0.919	0.984
Sanskrit	0.3483	0.867	0.988
Sinhala	0.7083	0.964	0.978
Spanish	0.5357	0.868	0.952
Tamil	0.4879	0.821	0.950
Telugu	0.3017	0.713	0.958
Thai	0.6173	0.856	0.919
Urdu	0.3385	0.831	0.978
Vietnamese	0.8913	0.981	0.988

Table 11. Performance comparison in terms of Average Precision (AP) at IoU [0.5–0.95] across 28 different languages over all document types in the UniTabBank dataset.

languages.

Statistical Reliability of Per-language Results: While per-language AP scores provide valuable insights, they can be misleading for very low-resource languages with extremely small samples (e.g., Farsi with only 18 images yielding AP=0.995). To ensure more reliable conclusions, we report average AP across three resource categories: High (500 samples), Medium (100–500 samples), and Low (<100 samples). As shown in Table 12, UniTabDet achieves strong performance across all categories, with AP=0.954 for High, AP=0.976 for Medium, and AP=0.993 for Low resource languages. The consistently high scores across categories demonstrate that, despite variability in sample size, UniTabBank supports robust and balanced performance across diverse linguistic settings.

F. More Visual Results

Success Example on Existing TD Benchmarks: Figs. 2–11 showcase representative examples where

Category	AP
All	0.959
Low	0.993
Medium	0.976
High	0.954

Table 12. Average Precision (AP) at IoU [0.5–0.95] for UniTabDet across different language resource categories (High, Medium, Low) and overall, demonstrating robust performance despite variations in sample availability.

UniTabDet successfully detects tables across multiple existing table detection benchmarks.

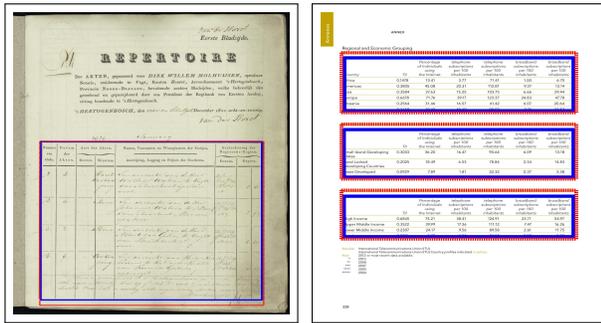


Figure 2. Shows visual results of ICDAR-2019 using UniTabDet. Red dotted box represents the table detections by UniTabDet while blue boxes indicate the ground truth.

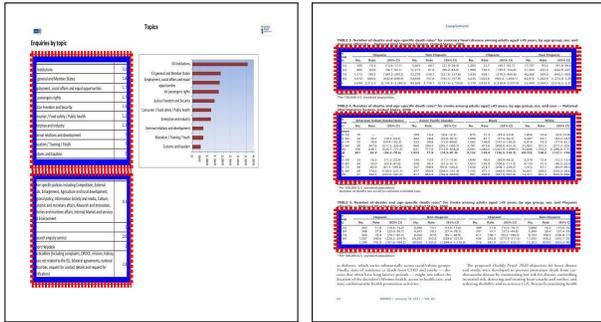


Figure 3. Shows visual results of ICDAR-2013 using UniTabDet. Red dotted box represents the table detections by UniTabDet while blue boxes indicate the ground truth.

Failure Example on Existing TD Benchmarks: Fig. 12 presents examples where UniTabDet appears to fail in table detection in existing table detection benchmarks. However, closer inspection reveals that the model predictions are correct, while the errors lie in the ground truth annotations.

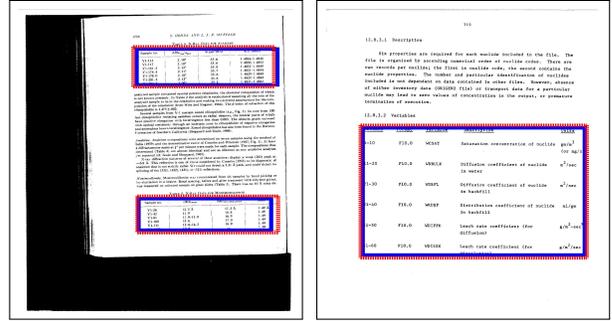


Figure 4. Shows visual results of UNLV using UniTabDet. Red dotted box represents the table detections by UniTabDet while blue boxes indicate the ground truth.

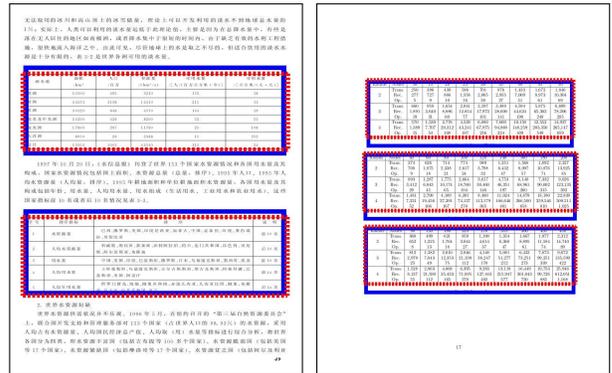


Figure 5. Shows visual results of Marmot using UniTabDet. Red dotted box represents the table detections by UniTabDet while blue boxes indicate the ground truth.

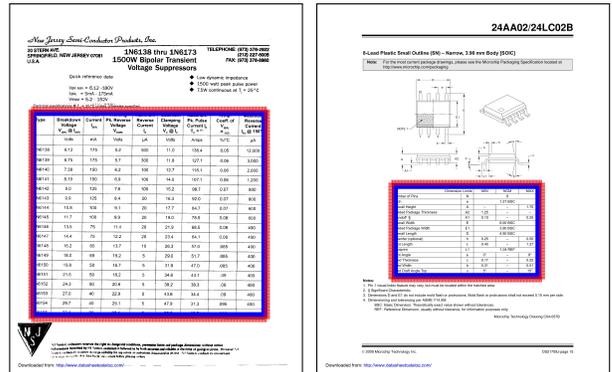


Figure 6. Shows visual results of STDW using UniTabDet. Red dotted box represents the table detections by UniTabDet while blue boxes indicate the ground truth.

G. Additional Experiments

We additionally evaluate UniTabDet in a zero-shot setting on four layout analysis benchmarks — *IIIT-AR-13K* [12], *DocLayNet* [14], *PubLayNet* [26], and *M⁶Doc* [3] — demonstrating its effectiveness in detecting tables across di-

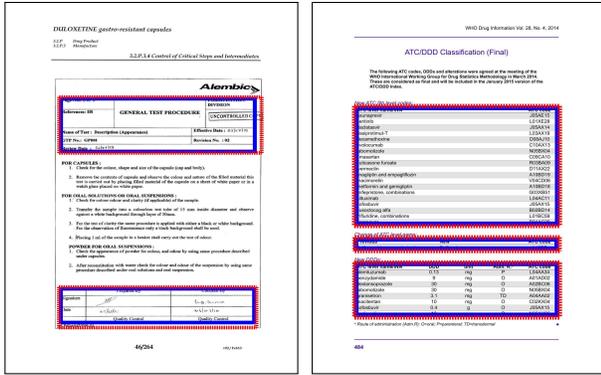


Figure 7. Shows visual results of **TNCR** using UniTabDet. Red dotted box represents the table detections by UniTabDet while blue boxes indicate the ground truth.

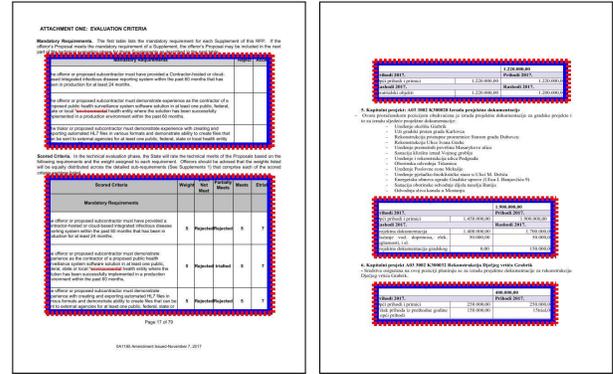


Figure 10. Shows visual results of **TableBank** using UniTabDet. Red dotted box represents the table detections by UniTabDet while blue boxes indicate the ground truth.

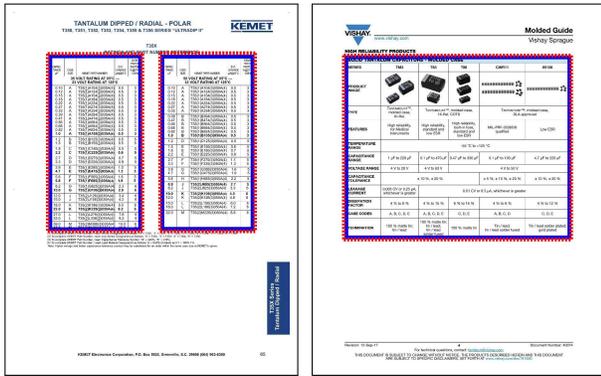


Figure 8. Shows visual results of **ICT-TD** using UniTabDet. Red dotted box represents the table detections by UniTabDet while blue boxes indicate the ground truth.

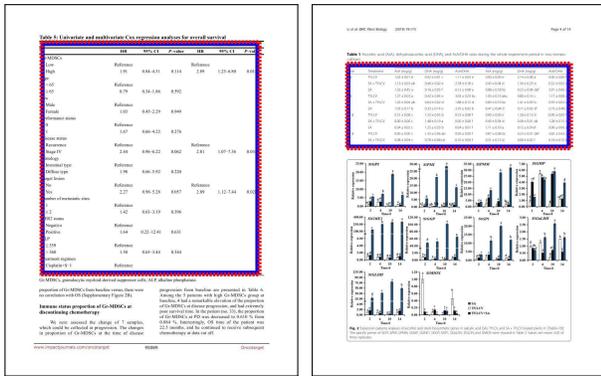


Figure 9. Shows visual results of **PubTables-1M** using UniTabDet. Red dotted box represents the table detections by UniTabDet while blue boxes indicate the ground truth.

Dataset	Model	AP: Document-Type
PubLayNet [26]	VGt [4]	0.981
	DETR [18]	0.981
	TRDLU [24]	0.976
	VSR [25]	0.974
	UniTabDet	0.923
DocLayNet [14]	YOLO [14]	0.863
	UniTabDet	0.839
M ⁰ Doc [3]	TransDLANet [3]	0.824
	UniTabDet	0.811
IIT-AR-13K [12]	Mask RCNN [12]	0.965
	UniTabDet	0.951

Table 13. Performance comparison of UniTabDet in **zero-shot** setting in terms of Average Precision (AP) at IoU [0.5–0.95] across four layout analysis benchmarks.

verse document types without additional fine-tuning.

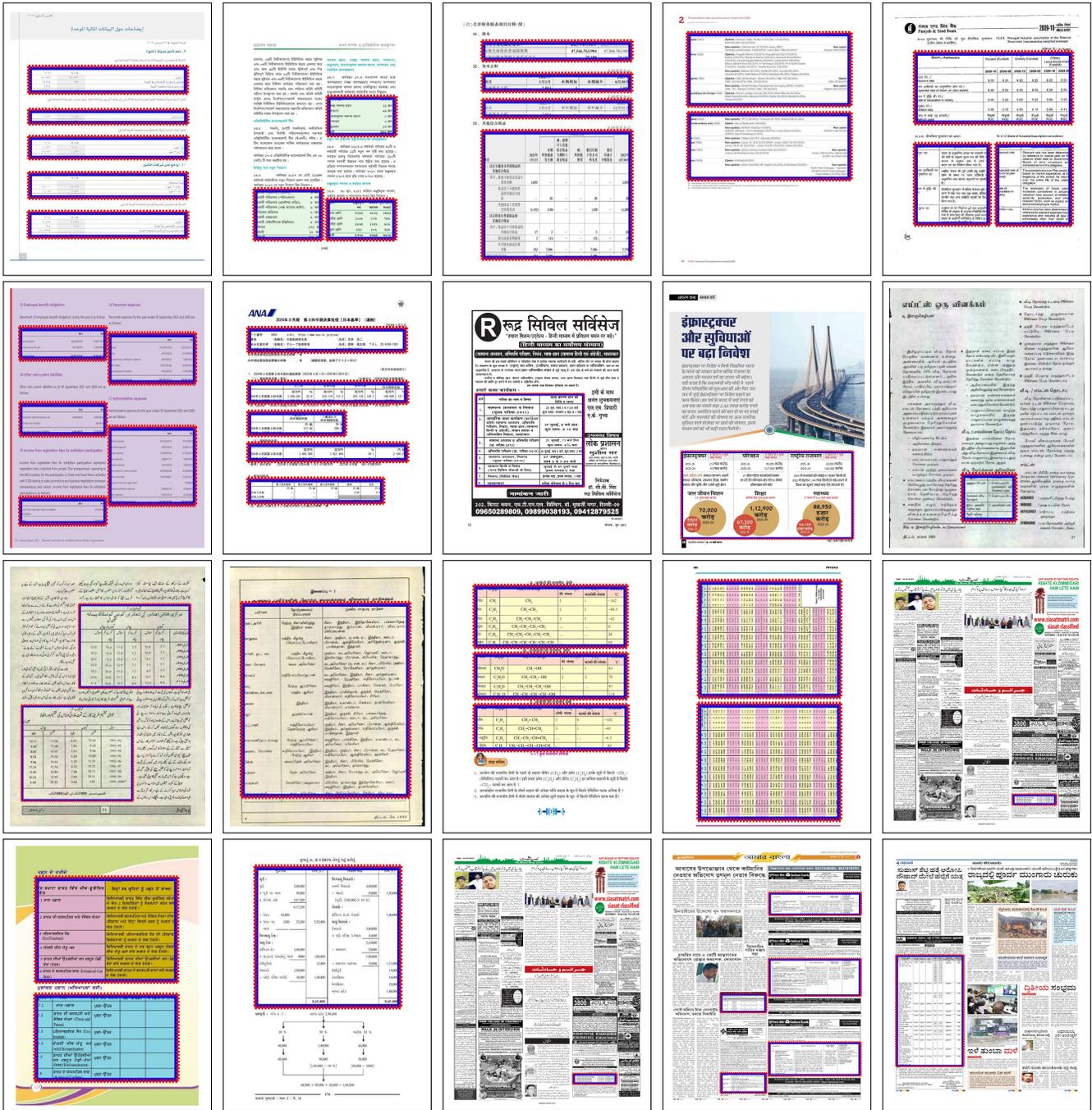


Figure 11. Shows visual results of UniTabBank using UniTabDet. Red dotted box represents the table detections by UniTabDet while blue boxes indicate the ground truth.

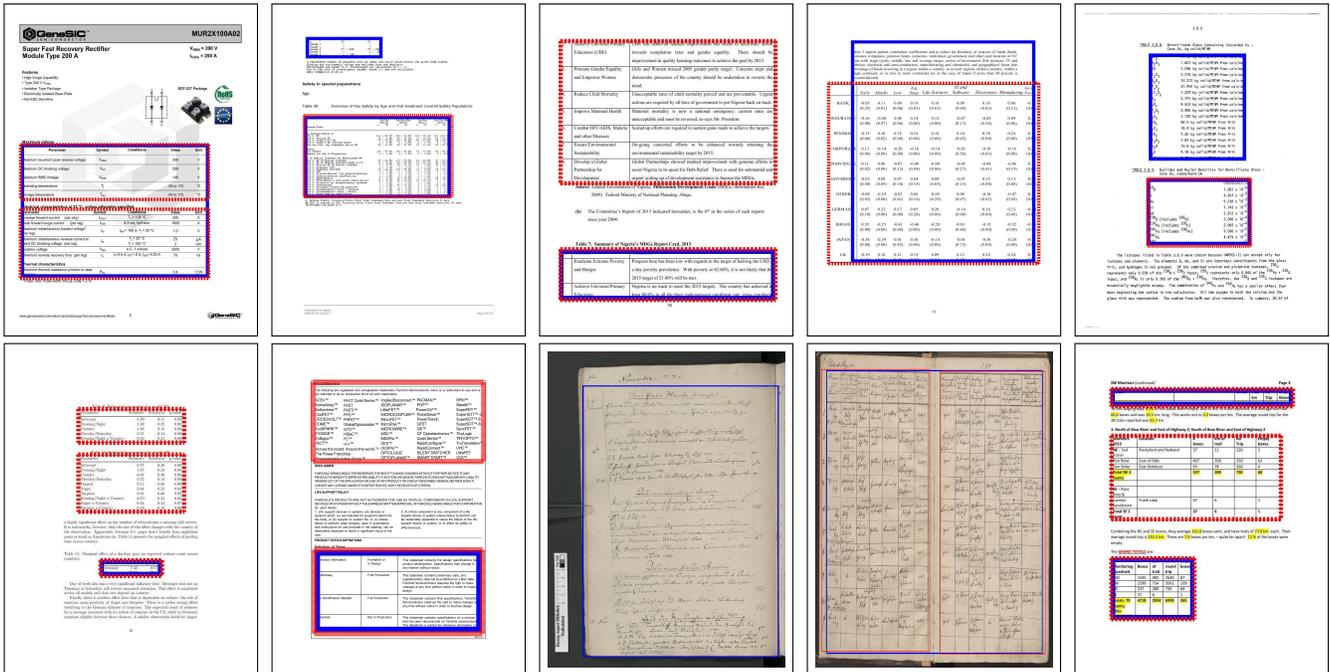


Figure 12. Shows a few failures and inaccurately detected examples from existing table detection benchmarks — ICDAR-2019, UNLV, STDW, ICT-TD, TNCR, TableBank, PubTables-1M using UniTabDet. The red dotted box represents the table detections by UniTabDet, while the blue boxes indicate the ground truth. In most cases, UniTabDet predicts tables that are not annotated in the ground truths. These failures occur due to inaccurate ground truths.

References

- [1] Abdelrahman Abdallah, Alexander Berendeyev, Islam Nuradin, and Daniyar Nurseitov. TNCR: Table net detection and classification dataset. *Neurocomputing*, 473:79–97, 2022. 4
- [2] Madhav Agarwal, Ajoy Mondal, and CV Jawahar. CDeC-Net: Composite deformable cascade network for table detection in document images. In *ICPR*, pages 9491–9498, 2021. 2, 4, 5
- [3] Hiuyi Cheng, Peirong Zhang, Sihang Wu, Jiaxin Zhang, Qiyuan Zhu, Zecheng Xie, Jing Li, Kai Ding, and Lianwen Jin. M6Doc: A large-scale multi-format, multi-type, multi-layout, multi-language, multi-annotation category dataset for modern document layout analysis. In *CVPR*, pages 15138–15147, 2023. 6, 7
- [4] Cheng Da, Chuwei Luo, Qi Zheng, and Cong Yao. Vision grid transformer for document layout analysis. In *CVPR*, pages 19462–19472, 2023. 7
- [5] Johan Fernandes, Murat Simsek, Burak Kantarci, and Shahzad Khan. TableDet: An end-to-end deep learning approach for table detection and table image classification in data sheet images. *Neurocomputing*, 468:317–334, 2022. 5
- [6] Liangcai Gao, Yilun Huang, Hervé Déjean, Jean-Luc Meunier, Qinqin Yan, Yu Fang, Florian Kleber, and Eva Lang. Icdar 2019 competition on table detection and recognition (ctdar). In *ICDAR*, pages 1510–1515, 2019. 2, 5
- [7] Mrinal Haloi, Shashank Shekhar, Nikhil Fandee, Siddhant Swaroop Dash, et al. Table detection in the wild: A novel diverse table detection dataset and method. *arXiv preprint arXiv:2209.09207*, 2022. 3, 4
- [8] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask R-CNN. In *ICCV*, pages 2961–2969, 2017. 3
- [9] Yilun Huang, Qinqin Yan, Yibo Li, Yifan Chen, Xiong Wang, Liangcai Gao, and Zhi Tang. A yolo-based table detection method. In *ICDAR*, pages 813–818, 2019. 5
- [10] Rahima Khanam and Muhammad Hussain. Yolov11: An overview of the key architectural enhancements. *arXiv preprint arXiv:2410.17725*, 2024. 3
- [11] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *ECCV*, pages 740–755, 2014. 3
- [12] Ajoy Mondal, Peter Lipps, and CV Jawahar. Iit-ar-13k: A new dataset for graphical object detection in documents. In *DAS*, pages 216–230, 2020. 6, 7
- [13] Shubham Singh Paliwal, D Vishwanath, Rohit Rahul, Monika Sharma, and Lovekesh Vig. TableNet: Deep learning model for end-to-end table detection and tabular data extraction from scanned document images. In *ICDAR*, pages 128–133, 2019. 4, 5
- [14] Birgit Pfitzmann, Christoph Auer, Michele Dolfi, Ahmed S Nassar, and Peter Staar. DocLayNet: A large human-annotated dataset for document-layout segmentation. In *ACM SIGKDD*, pages 3743–3751, 2022. 6, 7
- [15] Devashish Prasad, Ayan Gadpal, Kshitij Kapadni, Manish Visave, and Kavita Sultanpure. CascadeTabNet: An approach for end to end table detection and structure recognition from image-based documents. In *CVPRW*, pages 572–573, 2020. 4, 5
- [16] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. *NeurIPS*, 28, 2015. 3
- [17] Sebastian Schreiber, Stefan Agne, Ivo Wolf, Andreas Dengel, and Sheraz Ahmed. Deepdesrt: Deep learning for detection and structure recognition of tables in document images. In *ICDAR*, pages 1162–1167, 2017. 4, 5
- [18] Tahira Shehzadi, Khurram Azeem Hashmi, Didier Stricker, Marcus Liwicki, and Muhammad Zeshan Afzal. Bridging the performance gap between detr and r-cnn for graphical object detection in document images. *arXiv preprint arXiv:2306.13526*, 2023. 7
- [19] Shoaib Ahmed Siddiqui, Muhammad Imran Malik, Stefan Agne, Andreas Dengel, and Sheraz Ahmed. DeCNT: Deep deformable cnn for table detection. *IEEE access*, 6:74151–74161, 2018. 4, 5
- [20] B Smock, R Pesala, and R Abraham. Pubtables-1m: Towards comprehensive table extraction from unstructured documents. In *CVPR*, pages 4624–4632, 2021. 3, 4, 5
- [21] Yongzhou Wang, Wenliang Lv, Weijie Wu, Guanheng Xie, BiBo Lu, ChunYang Wang, Chao Zhan, and Baishun Su. TransTab: A transformer-based approach for table detection and tabular data extraction from scanned document images. *Machine Learning with Applications*, pages 100665–100678, 2025. 2, 4, 5
- [22] Bin Xiao, Murat Simsek, Burak Kantarci, and Ala Abu Alkheir. Table detection for visually rich document images. *Knowledge-Based Systems*, 282:111080–111115, 2023. 2, 3, 4, 5
- [23] Bin Xiao, Murat Simsek, Burak Kantarci, and Ala Abu Alkheir. Revisiting table detection datasets for visually rich documents. *IJDAR*, pages 1–20, 2025. 4
- [24] Huichen Yang and William Hsu. Transformer-based approach for document layout understanding. In *ICIP*, pages 4043–4047, 2022. 7
- [25] Peng Zhang, Can Li, Liang Qiao, Zhazhan Cheng, Shiliang Pu, Yi Niu, and Fei Wu. Vsr: a unified framework for document layout analysis combining vision, semantics and relations. In *ICDAR*, pages 115–130, 2021. 7
- [26] Xu Zhong, Jianbin Tang, and Antonio Jimeno Yepes. Publaynet: largest dataset ever for document layout analysis. In *ICDAR*, pages 1015–1022, 2019. 6, 7