A Lightweight Approach to Table Recognition in Digital Invoices

Pawel Marczak
Independed Researcher
Bialystok, Poland *moorless123@gmail.com*

Miroslaw Omieljanowicz
Bialystok University of Technology
Bialystok, Poland *m.omieljanowicz@pb.edu.pl*

Abstract

This study investigates table recognition techniques for digital documents, focusing on the challenges posed by diverse invoice layouts. A comparative evaluation of traditional pattern recognition and deep learning approaches highlighted their respective strengths and limitations. Special attention was given to ProjectionP, a proprietary lightweight method for resource-constrained environments, which combines morphological line extraction with pixel-based thresholding. A modified evaluation procedure adapted from ICDAR2013 was introduced to better balance over- and under-segmentation errors.

Comparative analysis showed that while Camelot leverages PDF metadata, it struggles with visual segmentation. Nanonets achieve high grid detection accuracy but can misplace text in complex tables. ProjectionP, optimized for desktop hardware, delivered competitive results, outperforming Camelot and matching Nanonets in specific cases.

Keywords: table recognition, table structure recognition, table segmentation, invoice processing, lightweight algorithm.

1. Introduction

Recent advances in deep image processing have greatly improved table structure recognition, enabling better handling of diverse formats. However, evaluating these methods remains challenging, requiring frameworks that balance accuracy with computational cost and data privacy. While deep learning dominates, traditional pattern recognition approaches still offer cost-effective alternatives, especially for SMEs. On datasets such as ICDAR-2013, both deep learning and traditional methods have achieved over 90% accuracy, showing that conventional techniques can remain effective.

Key evaluation criteria include accuracy, adaptability to enterprise data, ease of deployment, and confidentiality compliance. Deep learning models often require domain-specific fine-tuning, introducing overhead, whereas pattern recognition methods typically offer simpler implementation and lower data leakage risk.

This study presents a comparative evaluation of Camelot [1], NanoNets [11], and the proprietary ProjectionP method [10], focusing on their performance in recognizing tables within accounting documents and highlighting their strengths, limitations, and suitability for practical deployment.

2. Problem statement and current approaches

Accurate recognition of table structures in digital documents remains a complex challenge, particularly for invoices, forms, and reports. This has important implications for industries such as accounting, logistics, and data analytics, where automating tabular data extraction improves efficiency and reduces human error.

Traditional methods leverage geometric and morphological analysis. Profile projection techniques segment rows and columns by analyzing black pixel distributions [16], offering computational efficiency but limited flexibility. Clustering methods like DBSCAN group text blocks based on spatial patterns, which works well for structured tables but struggles with dense layouts [18]. Top-down approaches rely on geometric features such as frame detection and margin analysis [2], [7], while bottom-up methods detect individual elements and infer relationships [14], offering better adaptability at the cost of higher computational demands.

Deep learning approaches have advanced the field by using transformer-based models with self-attention mechanisms [17] and graph neural networks to capture complex structural relationships [13]. While highly accurate, these methods require significant computational resources and extensive training data.

Hybrid systems combine traditional and deep learning techniques, integrating OCR outputs with visual features [8] or using cascaded neural networks [4], aiming to balance accuracy, scalability, and resource efficiency in practical applications.

3. Proprietary Method

The recognition of table structures in digital documents is a critical challenge for small and very small enterprises, which often rely on desktop-class computers without access to cloud-based solutions due to concerns about data confidentiality. The proprietary method ProjectionP described in [10] addresses these challenges. It is specifically designed for implementation on standard PC-class hardware, enabling cost-effectiveand secure table structure recognition.

3.1. Core Concepts and Innovations

The method draws inspiration from Zuyev's algorithm [19], which uses projection-based inference for table segmentation. However, the new approach overcomes several limitations of traditional methods by introducing pixel-based projections and incorporating advanced preprocessing techniques [10].

The main concept involves the use of two complementary frame detection approaches:

- **Morphological Line Extraction**: Erosion with structured elements isolates horizontal and vertical lines, which are then projected onto their respective axes. Thresholding detects text boundaries;
- **Projection-Based Thresholding**: Raw pixel projections undergo normalization and exponential transformations to highlight separating lines, even in cases of interrupted frames or faint borders;

By merging the results from both approaches, the algorithm reduces false negatives and enhances frame detection robustness.

3.2. ProjectionP Method

The proprietary ProjectionP method [10] is central to the algorithm's ability to extract table structures efficiently. Its key steps include:

```
// Border Removal - Detected borders are removed
removeDetectedBorders(image)

// Text Enhancement - Dilation is applied, suppressing  noise
dilateImage(image, kernelSize=10x10)

// Signal Processing and Filtering - Generated signals are smoothed
signals = generateProjectionSignals(image)
smoothedSignals = applyCenteredMovingMaximumFilter(signals)
```

```
// Grid Refinement - Detected grid points from the frame detection
// and projection are reconciled
gridPoints = detectGridPoints(image)
refinedGrid = reconcileAndCleanGrid(gridPoints)

// Empty Row/Column Cleanup - Rows or columns that do
// not contain any text are eliminated
cleanedGrid = removeEmptyRowsAndColumns(refinedGrid, threshold)

// Cell Extraction - individual cells are extracted from the document
// Each cell's content is then processed separately
// to ensure accurate text capture
cells = extractCells(cleanedGrid, image)

// OCR Integration - The extracted cells are passed through an OCR module
// for each cell in cells:
text = performOCR(cell)
storeTextWithHierarchy(cell, text)

// Output Generation - The final data is organized
// into a structured CSV format
exportDataAsCSV(cells, outputPath)
```

4. Preparation of Data and Evaluation Method

Due to the fact that the problem of invoice processing was reported by Polish SMEs, the experiments were conducted using documents originating from Polish companies. A total of 267 PDF files from 12 different firms were collected. The dataset was divided into 12 styles, which were then used separately for comparing different methods. Using a free tool [9], annotations were prepared in VOC XML, YOLO, and CSV formats. The samples in the dataset varied in complexity. A general summary of the dataset is presented in Table 1.

Table 1. Table of Complexity and Border Types

	Fully bordered	Vertical bordered	Horizontal bordered
Simple Table	47	3	23
Partially Complex Table	11	19	0
Complex Table	0	0	9
Total	58	22	32

For comparative studies, two publicly available table extraction algorithms were selected. The first was Camelot [1], an open-source library for processing PDFs with embedded metadata. Despite this limitation, it was included due to the presence of suitable files in the dataset and its potential applicability. The second was Nanonets [11], a product widely adopted by large enterprises (e.g., AXA, Deloitte), selected as a quality benchmark. Results were obtained using Nanonets models publicly available at the end of 2023; newer versions may perform differently.

Selecting an evaluation method compatible with all solutions was challenging. Initially, widely referenced automated methods such as ICDAR2013 and ICDAR2019 were considered [5, 6]. ICDAR2019, though appropriate for computer vision approaches, was excluded because Camelot and Nanonets only export final CSV outputs without segmentation data. ICDAR2013 proved more feasible, relying on text comparison compatible with CSV files, but this approach disadvantages OCR-based methods and ignores empty cells.

A compromise between automation and reliability was adopted, allowing users to distinguish OCR errors from segmentation errors and to include empty cells. Although based on ICDAR2013 and requiring minimal verification, this approach still had limitations: frequent over- or under-segmentation along one axis led to low scores despite correct segmentation on the other. To address this, a less restrictive modification of ICDAR2013 was introduced. A dedicated tool was developed to manually tag cell connections in CSV files, enabling automated calculation of precision, recall, and F1-scores. The rules of the modified evaluation were as follows:

1. If text in a cell should have been in an adjacent cell or if segmentation errors were evident (e.g., clipped characters), the connection between the cells was classified as erroneous;
2. Connections between extra rows/columns caused by over- segmentation were classified as erroneous to prevent recall exceeding 1;
3. Improper merges of rows/columns were ignored, assuming the corresponding error was accounted for in the recall metric.

Consequently, our method is a less restrictive version of the original ICDAR2013. All connections classified as erroneous in the modification would also be erroneous in the original, but not vice versa.

5. Results and Discussions

The results obtained on the test set were evaluated in two stages. Initially, collective metrics were calculated across all 63 test tables by averaging and computing the standard deviations for precision, recall, and F1-score for the three compared methods: Camelot, Nanonets, and the author's method, referred to as ProjectionP (Table 2). Subsequently, the test set tables were grouped into 12 table styles (determined by the source document's provider), and the evaluation was repeated within these groups. This detailed analysis aimed to identify the strengths and weaknesses of each method (especially the author's) and to assess which methods are suitable for processing specific invoice categories in practical scenarios. The results achieved by the

Table 2. Comparison of Evaluation Metrics

	Camelot ICDAR	**Camelot Modified***	**ProjectionP ICDAR**	**ProjectionP Modified***	**Nanonets ICDAR**	**Nanonets Modified***
avg. precision	0.7346	0.8525	0.929	0.9787	0.9477	0.9849
std. precision	0.2843	0.2020	0.1491	0.0576	0.1020	0.0306
avg. recall	0.7860	0.9567	0.9063	0.9478	0.9400	0.9478
std. recall	0.2144	0.0608	0.1793	0.1003	0.1250	0.0645
avg. f1-score	0.7539	0.8906	0.9163	0.9661	0.9432	0.9785
std. f1-score	0.2590	0.1468	0.1665	0.0776	0.1141	0.0459

author's method appear promising when compared to the other methods. Using the ICDAR2013 evaluation methodology, ProjectionP's average F1-score of 0.91 is comparable to frequently cited solutions in the literature [12], [15]. However, these results are not directly comparable, as the evaluations were conducted on different datasets. The author's method is parameterized, while the DeepDeSRT [15] and TableNet [12] algorithms, operating on images, were likely tested with error penalties caused by OCR processing. Nevertheless, ProjectionP performed very well in the author's prepared comparison, outperforming the Camelot library by 0.16 in F1-score.

It is important to note that Camelot was tuned to prioritize recall over precision, often resulting in excessive segmentation. Adjusting parameters could improve precision at the expense of recall, potentially increasing the final F1-score slightly, but certainly not enough to match the author's method, which performed consistently better and more stably.

On the other hand, ProjectionP did not outperform Nanonets, which worked almost flawlessly in determining the table grid in all cases. Errors were primarily due to merged cells, which the method could not handle due to fundamental design assumptions.

The introduction of a modified evaluation method brought the results of the compared TSR algorithms closer together. Camelot particularly benefited from this change, achieving higher recall than the author's method (while its precision and overalln accuracy remained significantly weaker). Camelot's high recall (0.95) appears justified, as most actual text division points were detected, though excessive segmentation remained a persistent issue. From a practical application perspective, the high standard deviation values of the F1-scores for all three methods (exceeding 0.1) were concerning. Systems relying on these algorithms should be reliable, producing good, stable results. Below, we present results evaluated using the modified evaluation method.

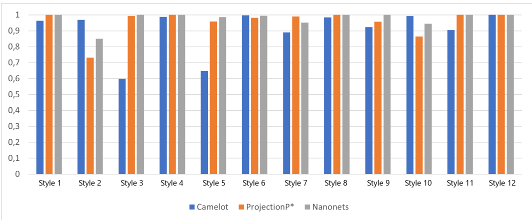

Fig. 1. Avarage F1-score results of methods evaluted with modified approach

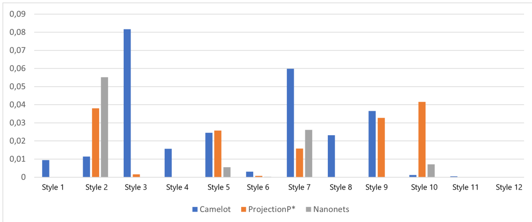

Fig. 2. Standard deviation of F1-score results of methods evaluted with modified approach

The F1-score metrics were recalculated, this time aggregating tables from source documents by provider (thus minimizing style variability within groups). As shown in Figures 1 and 2, each algorithm exhibited strengths and weaknesses depending on the table style. Unsurprisingly, Camelot was the most unreliable, completely failing for styles three and five (as represented in Figures 3 and 4). The poor results were caused by naive text line aggregation, which struggled with multi-line cells, resulting in poor row segmentation.

L.p.	Opis	Ilość	J.M.	Cena jednostk. przed rab.	Rab%	Cena jednostk. po rab.	Razem Netto	VAT
1	**Karton ozdobny Gładki kremowy 20 szt./op. 250 g/m2** Kod Indeksu: 202802	23	OP	4,90 PLN		4,90 PLN	112,70 PLN	23%
2	**Karton ozdobny Iceland diamentowa biel 20 szt./op. 220g/m2** Kod Indeksu: 200604	12	OP	5,90 PLN		5,90 PLN	70,80 PLN	23%
3	**Karton ozdobny Millenium błękitny 20 szt./op. 220g/m2** Kod Indeksu: 200708	12	OP	5,90 PLN		5,90 PLN	70,80 PLN	23%
4	**Folia laminacyjna PE arkusz A4 100 mic standard błysk antystatyczna** Kod Indeksu: 320410	5	OP	24,00 PLN		24,00 PLN	120,00 PLN	23%

Fig. 3. Example of a table in style three from the author's dataset

Lp	Symbol	EAN	Symbol obcy	Opis	JM	Ilość	Cena PLN	Netto PLN	Vat %	Vat PLN	Brutto PLN
20	S 175 COC12	4007817048481		Kredki sześciokątne, kolekcja Comic, 12 kol, Staedtler	szt	1	4,64	4,64	23%	1,07	5,71
21	S 965 14L NBK	4007817965023		Nożyczki Noris Club dla dzieci leworęcznych, 14 cm, blister, Staedtler	szt	2	3,63	7,26	23%	1,67	8,93
22	S 550 55 BK	4007817185667		Cyrkiel szkolny, z uniwersalnym adapterem do pisaków i ołówków, blister, Staedtler	szt	2	9,55	19,10	23%	4,39	23,49
23	MG HACT1257 BK	6941255141893		Korektor w taśmie So Many Cats, 2 szt, 5m x 5mm, MG	szt	10	4,53	45,30	23%	10,42	55,72
24	MG AS33V120	6941255143675		Karteczki samoprzylepne So Many Cats; 7,6x7,6cm, 80ark, MG	szt	20	2,38	47,60	23%	10,95	58,55
						Razem		571,94		131,57	703,51
								Netto	Vat %	Vat	Brutto
							W tym	571,94	23%	131,57	703,51

Fig. 4. Example of a table in style five from the author's dataset

Lp.	Kod Nazwa towaru	Kod EAN	Ilość/J.m.	Cena netto	Wartość netto	VAT
25	291 Papier (jk) A4-100 czerwony (fluo)	5905824300853	2 opak.	4,9700	9,94	23%
26	504 Pap.fluo.pomarańcz.samoprzylepny A4-20	5905824800599	1 opak.	8,7700	8,77	23%
27	501 Pap.fluo.żółty samoprzylepny A4-20	5905824800575	1 opak.	8,7700	8,77	23%
28	476 Papier srebrny samoprzylepny A4-20	5905824120628	2 opak.	14,5200	29,04	23%
29	648 Teczka (jkp)A4/10 szk.pastel.z gumką lakier	5905824010431	5 opak.	10,4000	52,00	23%
30	129 Terminarz INFO A6 (kal.ks.tyg.)	5905824600984	10 szt.	2,9100	29,10	23%

Fig. 5. Example of a table in style nine from the author's dataset

Interestingly, Camelot performed exceptionally well for style two (Figure 6), surpassing both ProjectionP and Nanonets. Its PDF-based approach, leveraging metadata in the file, cor-

Lp.	Symbol / Nazwa produktu		Ilość	J.m.	Rabat %	Cena netto	Kwota Netto	Podatek VAT %	Podatek VAT Kwota	Kwota Brutto
1	127321 FC	9556089009232	18	szt	20,00	2,15	38,70	P23	8,90	47,60
	GUMKA ARTYSTYCZNA CHLEBOWA MIX KOL.ETUIPLASTIKOWE CZERWONA/NIEBIESKA/ŻÓŁTA FABER-CASTELL [127321 FC]									
2	300/001/C ED	4004764390564	10	szt	20,00	2,07	20,70	P23	4,76	25,46
	MARKER EDDING PERMANENTNY OKRĄGŁA KOŃCÓWKA 1,5-3MM CZARNY [300/001/C ED]									
3	54410-21063 DA	4007885232898	10	szt		7,91	79,10	P23	18,19	97,29
	NOŻYCZKI DAHLE COMFORT-GRIP, DO PAPIERU24,7 cm [54410-21063 DA]									
4	54405-21092 DA	4007885232867	10	szt		3,23	32,30	P23	7,43	39,73
	NOŻYCZKI DAHLE COMFORT-GRIP, DOMOWE 14 cm [54405-21092 DA]									
5	117201 FC	4005401172017	12	szt	20,00	2,15	25,80	P23	5,93	31,73
	OŁÓWEK GRIP 2001 B Z GUMKĄ FABER-CASTELLCASTELL [117201 FC]									
6	114000 FC	4005401140009	1	kpl	20,00	17,59	17,59	P23	4,05	21,64
	OŁÓWKI GOLD FABER 1222 6 SZT. +GUMKA+TEMPERÓWKA OPAKOWANIE KARTONOWE FABER-CASTELL [114000 FC]									
7	59408-00000-00 TS	4042448366849	1	kpl	20,00	7,99	7,99	P23	1,84	9,83
	PŁATKI SAMOPRZYLEPNE TESA TACK 72 SZT. TRANSPARENTNE [59408-00000-00 TS]									

Fig. 6. Example of a table in style two from the author's dataset

Lp.	Column1	Symbol	Ilość	J.m.	Rabat	Cena
		Nazwa produktu			%	netto
1	127321 FC	9556089009232	18	szt	20,00	2,15
		GUMKA ARTYSTYCZNA CHLEBOWA MIX KOL.ETUIPLASTIKOWE CZERWONA/NIEBIESKA/ŻÓŁTA FABER-CASTELL [127321 F				
2	300/001/C ED	4004764390564	10	szt	20,00	2,07
		MARKER EDDING PERMANENTNY OKRĄGŁA KOŃCÓWKA 1,5-3MM CZARNY [300/001/C ED]				
3	54410-21063 DA	4007885232898	10	szt		7,91
		NOŻYCZKI DAHLE COMFORT-GRIP, DO PAPIERU24,7 cm [54410-21063 DA]				
4	54405-21092 DA	4007885232867	10	szt		3,23
		NOŻYCZKI DAHLE COMFORT-GRIP, DOMOWE 14 cm [54405-21092 DA]				
5	117201 FC	4005401172017	12	szt	20,00	2,15
		OŁÓWEK GRIP 2001 B Z GUMKĄ FABER-CASTELLCASTELL [117201 FC]				
6	114000 FC	4005401140009	1	kpl	20,00	17,59
		OŁÓWKI GOLD FABER 1222 6 SZT. +GUMKA+TEMPERÓWKA OPAKOWANIE KARTONOWE FABER-CASTELL [114000 FC]				
7	59408-00000-00 TS	4042448366849	1	kpl	20,00	7,99
		PŁATKI SAMOPRZYLEPNE TESA TACK 72 SZT. TRANSPARENTNE [59408-00000-00 TS]				

Fig. 7. Results of Camelot's operation for a table fragment from Figure 6

LP	Symbol towaru indeks odbiorcy/EAN	Nazwa towaru / (PKWiU)	Ilość J.M.	Cena jedn. bez podatku netto	Rabat %	Cena jedn. po rabacie	Wartość towaru/usł. po rabacie
1	**VLP-4100118** 5901478995032	PAPIER TOALET. VELVET, BIAŁY, 2W CELUL., OP. 8 ROL. PO 138 LISTKÓW (15 METRÓW) (17.22.11.0)	8 OP	3,47	0,00	3,47	27,76
2	**3M-UU00487408 5** 5902658076305	TAŚMA MONTAŻOWA SCOTCH, 19MMx1,5M, PRZEZROCZYSTA (22.29.21.0)	1 szt	11,80	0,00	11,80	11,80
3	**3M-UU00638003 2** 5900323008606	17001 PL HAKI COMMAND WIELOKROTNEGO UZYTKU, SREDNIE, 2 HAKI, 4 ŚREDNIE PASKI (22.29.22.0)	1 op	7,13	0,00	7,13	7,13
4	**3M-UU00638031 3** 5900323008385	17006 PL HACZYKI MAŁE, COMMAND, 6SZT, 8 MINI PASKÓW (22.29.22.0)	1 szt	7,13	0,00	7,13	7,13
5	**SKY-88031879** 7318761070006	PAPIER KSERO SKY COPY, A4, 500 ARKUSZY, KLASA C, GRAMATURA 80 (17.12.14.0)	120 op	8,67	0,00	8,67	1 040,40
						RAZEM	1 094,22
							1 094,22

Fig. 8. A fragment of a table representing the tenth style from the test dataset

LP	Symbol towaru ind	Nazwa towaru / (PKWiU)	Ilość J.M.	Cena j	Rabat 0/0	Cena je	Wartość
1	VLF-41001 18 5901478	PAPIER TOALET. VELVET, BIAŁY, 2w CELUL., (	8 OP	3,47	0	3,47	27,76
2	3M-UU00487408 5 59(	TAŚMA MONTAŻOWA SCOTCH, 19MMX1,5M	1 szt	11,8	0	11,80	11,8
3	3M-UU00638003 2 59(	17001 PL HAKI COMMAND WIELOKROTNEG(	1 OD	7,13	0	7,13	7,13
4	3M-UU00638031 3 59(	17006 PL HACZYKI MAŁE, COMMAND, GSZT, 8	1 szt	7,13	0	7,13	7,13
5	SKY-88031879 731876	PAPIER KSERO SKY COPY, A4, 500 ARKUSZY, K	120 Op	8,67	0	8,67	1040,4
						RAZEM	1094,22
	›_l,_l____ _"A' "A						1094,22
l... ll	ll.-nl..- . _1_:—'_n'						

Fig. 9. Results of ProjectionP's operation for a table fragment from Figure 8, with artifacts caused by the presence of text outside the table area

rectly grouped text into appropriate cells, even when it extended beyond the physical column boundaries (Figure 7).

ProjectionP generally performed well across other invoice styles, with the exception of style nine (Figure 5), where results exhibited a high deviation (over 0.1). Adjusting parameters (e.g. increasing the power exponent and reducing filter length) resolved these issues. With these changes, stability improved, and F1-score rose from 0.840 to 0.973 for this style, though minor errors (e.g., splitting words like "Code" and "VAT" in headers) persisted.

For style ten tables, another limitation was observed. When characters appeared outside the table area, ProjectionP included them in the output file, often fragmented due to segmentation (Figure 8). Although this did not affect segmentation itself, it introduced noise (Figure 9), underscoring the inadequacy of relying solely on rectangular region detection. For practical use, additional mechanisms should be implemented to remove such artifacts during pre- or post-processing.

While Nanonets appeared visually reliable (Figure 10), its OCR module sometimes misplaced or duplicated words near segmentation points (Figure 11). Although likely intended to avoid splitting words, this behavior was treated as a segmentation error in evaluation.

This issue contributed to ProjectionP outperforming Nanonets in styles seven and ten. The fact that each method excelled on specific styles motivated further analysis. For other invoice styles, method differences were minor and warranted additional detailed comparison.

6. Conclusions

The ProjectionP method proved highly effective on the proposed dataset. The algorithm performed well for certain styles and could be fine-tuned to avoid major errors in simple tables, although occasional header splitting required post-processing. However, due to diverse table layouts—particularly unbordered tables—this threshold-based approach cannot handle all cases. Similar limitations are noted in the literature [12], [13], [15], [19], and were observed for

lp.	mag	kod kreskowy	nazwa towaru/usługi
1	A01	5902277171221	BLOK BIUROWY A4 50# 70G
2	A01	5902277171238	BLOK BIUROWY A5 50# 70G
3	A01	5902277070005	BLOK TECHNICZNY A4 10 170G
4	A01	5902277171269	KOŁOZESZYT A4 80# M 70G Z PERF.UV
5	A01	5902277213044	TECZKA Z GUMKĄ A4+ CZERWONA
6	A01	5902277213082	TECZKA Z GUMKĄ A4+ POMARAŃCZOWA

Fig. 10. Segmentation results of the Nanonets method on a fragment of table in style seven from the author's dataset

lp.	mag	kodkreskowy	nazwatowaru/usługi
1	A01	5902277171221|BLOK	BIUROWY A4 50# 70G
2	A01	5902277171238|BLOK	BIUROWY A5 50# 70G
3	A01	A01|5902277070005	BLOK TECHNICZNY A4 10 170G
4	A01	5902277171269	KOŁOZESZYT A4 80# M 70G Z PERF.UV
5	A01	5902277213044|TECZKA	Z GUMKĄ A4+ CZERWONA
6	6|A01	5902277213082	TECZKA Z GUMKĄ A4+ POMARAŃCZOWA

Fig. 11. Final results (CSV file) of the Nanonets method's operation on a table fragment from Figure 10

Camelot and Nanonets.

While ProjectionP is practical for consistent table formats, as found in private enterprises with stable document types, it is less suitable for large-scale processing of heterogeneous data sources, where frequent parameter adjustments are needed. Technically, relying solely on profile projection is too limited for complex structures, though combining multiple segmentation techniques shows promise. However, this can also accumulate errors, such as introducing empty rows or columns.

Future work could focus on reducing the number of parameters or automating estimation using image-derived features [19] or unsupervised learning [18], though increased automation likely reduces accuracy. An end-to-end convolutional neural network approach inspired by TableNet [12], combined with the text-line classification module from ProjectionP, may yield superior results.

Automated evaluation remains an open challenge [3], [5], [18]. The modified ICDAR2013 evaluation better highlighted precision and recall differences and assessed partially correct tables less restrictively. Whether it is more useful than the original remains debatable; the choice depends on whether the priority is penalizing segmentation errors or differentiating between over- and under-segmentation. Automation of both evaluation methods could be possible using advanced text comparison metrics like BLEU [3], leveraging manually interpreted outputs as ground truth.

Acknowledgements The research was conducted under project no. WZ/I-IIT/5/2023 at Bialystok University of Technology and was financed through a research subsidy awarded by the minister responsible for science.

References

[1] Camelot Developers: Camelot: Pdf table extraction for humans. https://camelot-py.readthedocs.io/en/master/, accessed December 08, 2024

[2] Couasnon, B., Lemaitre, A.: Recognition of tables and forms. In: Handbook of Document Image Processing and Recognition, pp. 647–677. Springer, Cham, Switzerland (2014)

[3] Deng, Y., Rosenberg, D., Mann, G.: Challenges in end-to-end neural scientific table recognition. In: Proceedings of the International Conference on Document Analysis and Recognition (ICDAR). pp. 894–901 (2019)

[4] Devashish, P., Jaiswal, S., Patel, A.: Cascadetabnet: An approach for end-to-end table detection and structure recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 156–162 (2020)

[5] Déjean, H., et al.: Icdar 2019 competition on table detection and recognition (ctdar). https://doi.org/10.5281/zenodo.3239032, accessed November 08, 2024

[6] Göbel, M., et al.: Icdar 2013 table competition. https://roundtrippdf.com/en/data-extraction/dataset-format/, accessed November 08, 2024

[7] Huynh-Van, T., et al.: Learning to detect tables in document images using line and text information. In: Proceedings of the ACM SIGIR Conference (2018)

[8] Lin, C., Wang, X., Zhang, H.: Multimodal-tsr: Combining ocr and deep features for table structure recognition. International Journal on Document Analysis and Recognition 26, pp. 145–158 (2023)

[9] MakeSense AI: Makesense: Annotation tool for machine learning. https://www.makesense.ai/, accessed December 08, 2024

[10] Marczak, P., Omieljanowicz, M.: A new method projectionp for table structure recognition. In: Proceedings of the 23rd International Conference on Computer Information Systems and Industrial Management (CISIM). vol. 14902, pp. 74–88. Lecture Notes in Computer Science, Białystok, Poland (2024)

[11] Nanonets: Overview of api documentation. https://docs.nanonets.com/reference/overview, accessed November 18, 2023

[12] Paliwal, S.S., et al.: Tablenet: Deep learning model for end-to-end table detection and tabular data extraction from scanned document images. In: Proceedings of the International Conference on Document Analysis and Recognition (ICDAR). pp. 203–210 (2019)

[13] Prasad, D., Patel, A., Jaiswal, S.: Graphtabnet: Enhancing table recognition using graph neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 145–158 (2023)

[14] Raja, S., Mondal, A., Jawahar, C.V.: Table structure recognition using top-down and bottom-up cues. In: Proceedings of ECCV. pp. 70–86. Springer, Cham, Switzerland (2020)

[15] Schreiber, S., et al.: Deepdesrt: Deep learning for detection and structure recognition of tables in document images. In: Proceedings of ICDAR (2017)

[16] Sun, X., Liu, J., Wang, Z.: Projection profile-based table detection for historical documents. Pattern Recognition Letters 138, pp. 118–127 (2020)

[17] Zeng, W., Yu, J., Huang, Y.: Tableformer: A transformer-based framework for table detection and parsing. IEEE Transactions on Multimedia 24, pp. 1–12 (2022)

[18] Zucker, A., et al.: Clusti: Clustering method for table structure recognition in scanned images. Multimedia Tools and Applications 26, pp. 1765–1776 (2021)

[19] Zuyev, K.: Table image segmentation. In: Proceedings of the IEEE International Conference on Document Analysis and Recognition (ICDAR). pp. 705–708 (1997)