TableFormer: Table Structure Understanding with Transformers Supplementary Material

Ahmed Nassar, Nikolaos Livathinos, Maksym Lysak, Peter Staar IBM Research

{ahn,nli,mly,taa}@zurich.ibm.com

1. Details on the datasets

1.1. Data preparation

As a first step of our data preparation process, we have calculated statistics over the datasets across the following dimensions: (1) table size measured in the number of rows and columns, (2) complexity of the table, (3) strictness of the provided HTML structure and (4) completeness (i.e. no omitted bounding boxes). A table is considered to be simple if it does not contain row spans or column spans. Additionally, a table has a strict HTML structure if every row has the same number of columns after taking into account any row or column spans. Therefore a strict HTML structure looks always rectangular. However, HTML is a lenient encoding format, i.e. tables with rows of different sizes might still be regarded as correct due to implicit display rules. These implicit rules leave room for ambiguity, which we want to avoid. As such, we prefer to have "strict" tables, i.e. tables where every row has exactly the same length.

We have developed a technique that tries to derive a missing bounding box out of its neighbors. As a first step, we use the annotation data to generate the most fine-grained grid that covers the table structure. In case of strict HTML tables, all grid squares are associated with some table cell and in the presence of table spans a cell extends across multiple grid squares. When enough bounding boxes are known for a rectangular table, it is possible to compute the geometrical border lines between the grid rows and columns. Eventually this information is used to generate the missing bounding boxes. Additionally, the existence of unused grid squares indicates that the table rows have unequal number of columns and the overall structure is non-strict. The generation of missing bounding boxes for non-strict HTML tables is ambiguous and therefore quite challenging. Thus, we have decided to simply discard those tables. In case of PubTabNet we have computed missing bounding boxes for 48% of the simple and 69% of the complex tables. Regarding FinTabNet, 68% of the simple and 98% of the complex tables require the generation of bounding boxes.

Figure 1 illustrates the distribution of the tables across

different dimensions per dataset.

1.2. Synthetic datasets

Aiming to train and evaluate our models in a broader spectrum of table data we have synthesized four types of datasets. Each one contains tables with different appearances in regard to their size, structure, style and content. Every synthetic dataset contains 150k examples, summing up to 600k synthetic examples. All datasets are divided into Train, Test and Val splits (80%, 10%, 10%).

The process of generating a synthetic dataset can be decomposed into the following steps:

1. Prepare styling and content templates: The styling templates have been manually designed and organized into groups of scope specific appearances (e.g. financial data, marketing data, etc.) Additionally, we have prepared curated collections of content templates by extracting the most frequently used terms out of non-synthetic datasets (e.g. PubTabNet, FinTabNet, etc.).

2. Generate table structures: The structure of each synthetic dataset assumes a horizontal table header which potentially spans over multiple rows and a table body that may contain a combination of row spans and column spans. However, spans are not allowed to cross the header - body boundary. The table structure is described by the parameters: Total number of table rows and columns, number of header rows, type of spans (header only spans, row only spans, column only spans, both row and column spans), maximum span size and the ratio of the table area covered by spans.

3. Generate content: Based on the dataset *theme*, a set of suitable content templates is chosen first. Then, this content can be combined with purely random text to produce the synthetic content.

4. Apply styling templates: Depending on the domain of the synthetic dataset, a set of styling templates is first manually selected. Then, a style is randomly selected to format the appearance of the synthesized table.

5. Render the complete tables: The synthetic table is finally rendered by a web browser engine to generate the



Figure 1: Distribution of the tables across different dimensions per dataset. Simple vs complex tables per dataset and split, strict vs non strict html structures per dataset and table complexity, missing bboxes per dataset and table complexity.

bounding boxes for each table cell. A batching technique is utilized to optimize the runtime overhead of the rendering process.

2. Prediction post-processing for PDF documents

Although TableFormer can predict the table structure and the bounding boxes for tables recognized inside PDF documents, this is not enough when a full reconstruction of the original table is required. This happens mainly due the following reasons:

- TableFormer output does not include the table cell content.
- There are occasional inaccuracies in the predictions of the bounding boxes.

However, it is possible to mitigate those limitations by combining the TableFormer predictions with the information already present inside a programmatic PDF document. More specifically, PDF documents can be seen as a sequence of PDF cells where each cell is described by its content and bounding box. If we are able to associate the PDF cells with the predicted table cells, we can directly link the PDF cell content to the table cell structure and use the PDF bounding boxes to correct misalignments in the predicted table cell bounding boxes.

Here is a step-by-step description of the prediction postprocessing:

1. Get the minimal grid dimensions - number of rows and columns for the predicted table structure. This represents the most granular grid for the underlying table structure.

2. Generate pair-wise matches between the bounding boxes of the PDF cells and the predicted cells. The Intersection Over Union (IOU) metric is used to evaluate the quality of the matches.

3. Use a carefully selected IOU threshold to designate the matches as "good" ones and "bad" ones.

3.a. If all IOU scores in a column are below the threshold, discard all predictions (structure and bounding boxes) for that column. 4. Find the best-fitting content alignment for the predicted cells with good IOU per each column. The alignment of the column can be identified by the following formula:

$$alignment = \arg\min_{c} \{D_{c}\}$$

$$D_{c} = max\{x_{c}\} - min\{x_{c}\}$$
(1)

where c is one of {left, centroid, right} and x_c is the x-coordinate for the corresponding point.

5. Use the alignment computed in step 4, to compute the median x-coordinate for all table columns and the median cell size for all table cells. The usage of median during the computations, helps to eliminate outliers caused by occasional column spans which are usually wider than the normal.

6. Snap all cells with bad IOU to their corresponding median *x*-coordinates and cell sizes.

7. Generate a new set of pair-wise matches between the corrected bounding boxes and PDF cells. This time use a modified version of the IOU metric, where the area of the intersection between the predicted and PDF cells is divided by the PDF cell area. In case there are multiple matches for the same PDF cell, the prediction with the higher score is preferred. This covers the cases where the PDF cells are smaller than the area of predicted or corrected prediction cells.

8. In some rare occasions, we have noticed that Table-Former can confuse a single column as two. When the postprocessing steps are applied, this results with two predicted columns pointing to the same PDF column. In such case we must de-duplicate the columns according to highest total column intersection score.

9. Pick up the remaining orphan cells. There could be cases, when after applying all the previous post-processing steps, some PDF cells could still remain without any match to predicted cells. However, it is still possible to deduce the correct matching for an orphan PDF cell by mapping its bounding box on the geometry of the grid. This mapping decides if the content of the orphan cell will be appended to an already matched table cell, or a new table cell should be created to match with the orphan.

9a. Compute the top and bottom boundary of the horizontal band for each grid row (min/max y coordinates per row).

9b. Intersect the orphan's bounding box with the row bands, and map the cell to the closest grid row.

9c. Compute the left and right boundary of the vertical band for each grid column (min/max x coordinates per column).

9d. Intersect the orphan's bounding box with the column bands, and map the cell to the closest grid column.

9e. If the table cell under the identified row and column is not empty, extend its content with the content of the orphan cell.

9f. Otherwise create a new structural cell and match it wit the orphan cell.

Aditional images with examples of TableFormer predictions and post-processing can be found below.

Variable	Odds ratio	95% confidence interval	p value
ajor vascular complications	9 .91	9 .67–9.14	0.0 13
Renal failure required CRRT	12.98	3 21–19.41	40 .001
Severe bleeding	15.86	3.61-69.63	40 .001
Meurologic complications	20.68	288-34.80	22 0.001
		confidence interval	
Major vascular complications	5 .9	\$.67–9.14	0.013
Renal failure required CRRT	\$0.98	£0 21–19.41	10 .001
Severe bleeding	13 .86	3461-69.63	15 .001
Neurologic complications	12.68	18 8-34.80	19 .001
Post-processed bounding	boxes	8504	r value
Mariable	Odds ratio	confidence interval	Harditie

TableFormer predicted structure

anal failure required

🗯 vere bleeding

0 (0, 0	10	1. 0]	2 [2, 0]		3 (3,
Variable	Odds ratio	95% confide	ence	p value	
		interval			
4 (0, 1)	5 (. 1]	6 [2, 1]		7 (3,
Major vascular	3.91	1.67-9.14		0.013	
complications					
8 [0, 2]	9 (1, 2)	10 [2, 2		11 p.
Renal failure	10.98	6.21-19.41		< 0.001	
required CRRT					
12 [0, 3]	13 (1, 3]	14 [2, 3]		15 (3,
Severe bleeding	15.86	3.61-69.63		< 0.001	
16 [0, 4]	17 (1, 4]	18 [2, 4]		19 (3,
Neurologic	13.68	5.38-34.80		< 0.001	
complications					

10.98

15.86

25.68

121

3.61-69.63

288-34.80

40.001

40.001

450.001

Figure 2: Example of a table with multi-line header.

PDF Cells	
Name	Sequences
KRAS=F	5-IGIGIGACAIGTICIAAIAIAGICACAITI-3
KRAS R	5'-ATCGTCAAGGCACTCTTGCCTAC-3
PNA clamp probe	5-TACGCCACCAGCTCC-3
TableFormer predicted bounding bo	xes
Name	*Sequences
KRAS-F	5'-TGTGTGACATGTTCTAATATAGTCACATTT-3
KRAS-R	5'-ATCGTCAAGGCACTCTTGCCTAC-3'
PNA clamp probe	5-TACGCCACCAGCTCC-3
Post-processed bounding boxes	
Name	Sequences
KRAN	9-TGTGTGACATGTTCTAATATAGTCACATTT-3
KRAN®R	3-ATCGTCAAGGCACTCTTGCCTAC-3
PNA clamp probe	D-TACGCCACCAGCTCC-3

TableFormer pred	licted structure	
0 (0, 0		10
Name	Sequences	
2 (0, 1		3 (1
KRAS-F	5'-TGTGTGACATGTTCTAATATAGTCACATTT	ſ-3'
4 (0, 2		5 (1
KRAS-R	5'-ATCGTCAAGGCACTCTTGCCTAC-3'	
6 (0, 3		71
PNA clamp probe	5'-TACGCCACCAGCTCC-3'	

Figure 3: Example of a table with big empty distance between cells.

		ANO	VA	
	Sum Sq	Df	⊮ V alue	₽r (>F)
P	3745.2	1	266.75	4964 14 18 334
and	24 91.39	2	50.87	2976 20 20 2223
B a conc	2648.33	2	61.48	\$0()7 as \$2 3334
Residuals	286.91	M	B ∎	ef]
ableFormer pi	redicted bour	nding box	es	
		ANO	VA	2
3	Sum Sq	۶Df	F Value	Pr (>F)
P	\$745.2	10	266.75	4.64×10^{-9}
vonc	2191.39	2	50.87	276×10^{-6}
Pas conc	2648.33	25	61.48	1.07×10^{-6}
Residuals	236.91	351	36	20
ost-processe	d bounding b	oxes ANO	VA	
ost-processe	d bounding b	oxes ANO	VA MValue	¥r (>F)
ost-processe	d bounding b Sum Sq 8745.2	oxes ANO PI	VA Value	<u>₽т (>F)</u>
ost-processed	d bounding b Sum Sq 6745.2 491.39	oxes ANO DI DI	VA (Value) 266.75 50.87	<mark>۲۲ (>۲)</mark> <u>4864 און 1817</u> 1876 און 1812
est-processed	d bounding b Sum Sq 6745.2 6491.39 2548.33	oxes ANO DI B D D D	VA 10 Value 166.75 100.87 161.48	F T (>F) <u>1</u>454 (a) (1917) 1275 (a) (1917) 1907 (a) (1917) 1907 (a) (1917)

Figure 4: Example of a complex table with empty cells.

PDF Cells			
	3 mM (ng/ml/islet)	16.7 mM (ng/ml/islet)	Fold-increas (high GLC/ low GLC)
86 c <mark>ontro]</mark> 86 v <mark>ild</mark> agliptin treated 86(A) ⁸ [20ntro] 86(A) ⁸ [6]dagliptin treated	0.047 ±10.07 0.048 ±0.006 0.054 ±10.004 0.050 ±0.007	3396 140327 4234 23 24 24 2 45 8406 25 3407 47 1481 37 9601	0.05.3 9.86.3 27 3.86.3 29 (994.2 1
TableFormer predicted	bounding box	(es	
	3 mM (ng/ml/islet)	06.7 mM (ng/ml/islet)	Fold-increas (high GLC/ lcw GLC)
186 control 186 vildagliptin treated 186KA? control 18KA? vildagliptin treated	0.247 ± 0.07 10:248 ± 0.06 12:34 ± 0.04 12:30 ± 0.07	2.96±0.27 [4834±0.324 [2206±0.074] [2k.81±0.30↑]	6263 93439 93439 63427
Post-processed boundi	ng boxes		
	3 mM (.ag/ml/islet)	16.7 mM (ng/ml/islet)	Fold-increas (high GLC/ Low GLC)
46 cientro 56 vialagliptin treated 26 Ajabantro 46 Abbaldagliptin treated	3047(±1130)7 3048(±0420)6 (553-(±1420)6 (553-(±1420)6	10/6(100027) 405-4(10027) 1006(1007)\$7 1081(10000000	5253 5243357 5243359 5243359 5243359 5243359

TableFormer predicted structure			
0* p. e	1" [1.0	2* p. e	3 p. a
			Fold-increase
4* p. t	5(0.1)	612.1	7 p. ŋ
	3mM	16.7mM	(high GLC/
8* (p, z	9 [1, 2]	10 p. z	11 p. z
	(ng/ml/islet)	(ng/ml/islet)	low GLC)
12 p. s	13 (1.3	14 (2, 3	15 (p. s)
B6control	0.47±0.07	2.96±0.27	6.63
16 p. 4	17 (1.4	18 p. 4	19 p. 4)
B6vildagliptin treated	0.48±0.06	4.34±0.32*	9.43+
20 p. s	21 (1.8)	22 (2. 8	23 (5. 6)
KKAycontrol	0.34±0.04	1.06±0.07+	3.43+
24 p. s	26 (1. 0	26 p. s	27 (5.4)
KKAyvildagliptin treated	0.30±0.07	1.81±0.30†	6.42†

Figure 5: Simple table with different style and empty cells.

PDF Cells						
freatment	fank number	FCO ₂ (patm)	PH	fotal alkalinity (emol kg:4) ³²	Selinity (ppt)	Hemperature (°C)
Control	•	397 107 5	896 2006	2145 ±47.7	35.6 ±0.07	28.6 20.05
Control	p.	384 100. 8	8718 00.006	2145 a#7	55.6 ± 0.07	28.4 90.04
Medium	ľ	614 3216.6	5.00 ± 0.009	2095 x 9.1	55.9 i 87.07	28.7 x \$5.05
Medium	By	508 to 76.5	8:00 20.009	2095 209.1	35.9 20.07	28.6 # 2.05
High	6	876 144.6	786 20.006	2079 #2.3	36.0 10.03	28.7 alt.03
LINE IN	The state	182	207 327000	1070 - 310	11101-1110	1147 3 318 4

TableFormer predicted bounding boxes

Treatment	Tank number	PCO ₂ (µatm)	₽н	Total alkalinity (µmol kg ⁻¹)	Salinity (ppt)	Temperature (°C)
Control	00	397±6.5	8016±0.006	£145±4.7	135.6±0.07	28.6 ± 0.05
Gontrol	12	384±6.8	8:18±0.006	12145±4.7	185.6±0.07	26.4±0.04
Medium	國	2614±16.6	[8400±0.009]	12095 ± 5.1	285.9 ± 0.07	28.7 ± 0.05
Medium	23	#08±16.5	8:00±0.009	2095 ± 5.1	335.9±0.07	28.6±0.05
bligh		:876±14.6	[3 .86±0.006]	2079±5.3	486.0±0.03	41 28.7 ± 0.03
High	42	461±14.4	47.87 ± 0.006	42079±5.3	4736.0 ± 0.03	a28.7 ± 0.04

t-processed bounding boxes

ireatment	lank number	FCO ₂ (patm)	pH	iotal alkalinity	Salinity (ppt)	Hemperature (°C
Control	6	397 300.5	816 30.006	2145 国際7	35.6 30.07	28.6 30.05
26ntro	g.	384 30.8	8:18 0.006	2145 147	35.6	28.4 30.04
Wedium	.	014 1406.6	8:00 3 8.009	2095 389.1	35.9 H Ø.07	28.7 38.05
Medium	8	508 197 6.5	8.00 10.009	2095 377.1	35.9 30.07	28.6 10.05
ffigh	0*	876 3 44.6	7.86 .006	2879 38.3	38.0 10.03	28.7 30.03
ING I-	7	112 11 and /	1157 34704	11970 33310	1220 134800	110715000

TableForm	er pre	dicted structure										
	0 (0, 0	1 (1.0		2 (2, 4)		3 (a, c)		4 (4, 0		5 (5, 0)		6 (8,
Treatment		Tank number	PCO\$_{2}\$ (µatm)		рН		Total alkalin (µmol kg-1)	ity	Salinity (ppt)	Temperature (*C)	3
	7 (0, 1	8 (1, 1)		9 (2, 1		10 (J. 1)		11 (6, 1)		12 (5, 1)		13 (5
Control		1	397 ± 6.5		8.16 ± 0.006		2145 ± 4.7		35.6 ± 0.07		28.6 ± 0.05	
	14 (0, 2	15 (1. 2		16 (2, 2		17 (p. z)		18 (4, 2)		19 (5.2)		20 (A.
Control		2	384 ± 6.8		8.18 ± 0.006		2145 ± 4.7		35.6 ± 0.07		28.4 ± 0.04	
	21 (0.3	22 (1.3)		23 (2, 3)		24 (p. 3)		25 (4, 3)		26 (5.3)		27 (5
Medium		1	614 ± 16.6		8.00 ± 0.009		2095 ± 5.1		35.9 ± 0.07		28.7 ± 0.05	
	28 (0.4	29 (1.4		30 (2.4)		31 (p. 4)		32 (4, 4)		33 (5.4)		34 (8,
Medium		2	608 ± 16.5		8.00 ± 0.009		2095 ± 5.1		35.9 ± 0.07		28.6 ± 0.05	
	35 jp, s	36 (1, 5)		37 (2, 5		38 (a, s)		39 (4, 5)		40 (5, 5)		41 (8,
High		1	876 ± 14.6		7.86 ± 0.006		2079 ± 5.3		36.0 ± 0.03		28.7 ± 0.03	
	42 (0,6	43 (1.4)		44 (2.4)		45 p. e		46 (4, 6)		47 (5.6)		45 (8.
High		2	861 ± 14.4		7.87 ± 0.006		2079 ± 5.3		36.0 ± 0.03		28.7 ± 0.04	

Figure 7: Table predictions example on colorful table.

Variable	Sensitivity (%)	Specificity (%)	Cut off
Fotal Bilirubin	60	<mark>9</mark> 5	₽,3 mg/dl
Direct Bilirubin	60	95	0,85
🚱 Reactive Protein	47	85	96
ableFormer predicted bounding b	oxes		
∛ariable	Sensitivity (%)	Specificity (%)	Cut off
Total Bilirubin	160	95	[1,3 mg/dl]
Direct Bilirubin	\$60	95	49,85
C- Reactive Protein	14[7]	86	198
ost-processed bounding boxes			
variable	Sensitivity (%)	Specificity (%)	E ut off
Fotal Bilirubin	Đ	95	k,3 mg/dl
Direct Bilirubin	60	90	9 85
	871	851	881

·····			
0 (0.0)	1 (1, 0)	2 (2.0	3 (1.0
Variable	Sensitivity (%)	Specificity (%)	Cut off
4 (0, 1)	5 (1, 1)	6 (2, 1	7 p.
Total Bilirubin	60	95	1,3 mg/dl
8 (0, 2)	9 (1, 2)	10 (p. z	11 p.:
Direct Bilirubin	60	95	0,85
12 (2, 3)	13 (1, 3)	14 (2, 3	15 p.:
C- Reactive Protein	47	85	98

Figure 6: Simple table predictions and post processing.

PDF Cells Gortical Layer Grade 4 Grade 3 Grade 2 Grade 1 wniformly thick and cellular Molecular Sanable the Variable thinning and meduced cellularity niformly thin barge gaps and conspicuously in pouronal necrosis Rankinje Moderate gaps and mattered loss of neu densely cellular laregular thinning with modest reductions in cell biniformly thick and enspicuous reduction

ableFormer predicted bounding boxes

•Cortical Layer	Grade 4	*G rade 3	Grade 2	Grade 1
Molecular	 Uniformly thick and cellular 	Variable thinning, normal cellularity	*Variable thinning and reduced cellularity	Uniformly thin
Purkinje	Well populated with Histologically intact pyramidal neurons	alsolated neuronal loss or eosinophilic degeneration (necrosis)	Moderate gaps and scattered loss of neurons	*Large gaps and conspicuously increased neuronal necrosis
Granule	Uniformly thick and densely cellular	Mregular thinning but densely cellular	Irregular thinning with modest reductions in cell density	¹³ Irregular thinning and conspicuous reductions in cell density

1	Post-processed bounding boxes							
	o ortical Layer	Grade 4	Grade 3	6 rade 2	State sanformiv tan sanspecuously increased seuronal necross			
	Molecular	eniformly thick and reliatan	variable thinning, normal selfularity	sariable thinning and reduced cellularity	Saniformiy thin			
	farkinje	well populated with histologically intact nyramidal neurons	solated neuronal loss or misinophilic degeneration mecrosis	bioderate gaps and mattered loss of neurons	narge gaps and narge gap and			
	Geranule	conselv cellular	megular thinning but Innselv cellular	megular thinning with modest reductions in cell density	pregular thinning and j penspicuous reductions in pell density			

TableFormer predicted structure

	0 (0, 0)	1 (1.0	2 p. 0	3 p. o	4 (A. Q
	Cortical Layer	Grade 4	Grade 3	Grade 2	Grade 1
	5 (0, 1)	6 (1, 1	7 (2, 1	8 p. t	9 (6.1)
	Molecular	Uniformly thick and cellular	Variable thinning, normal	Variable thinning and	Uniformly thin
			cellularity	reduced cellularity	,
	10 (0, 2)	11 (1.2)	12 p. z	13 p. 2	14 p. g
	Purkinje	Well populated with	Isolated neuronal loss or	Moderate gaps and	Large gaps and
		histologically intact	eosinophilic degeneration	scattered loss of neurons	conspicuously increased
		pyramidal neurons	(necrosis)		neuronal necrosis
	15 (p. a)	16 (r. a	17 (2.3	18 p. a	19 (4. 3)
	Granule	Uniformly thick and densely	Irregular thinning but densely	Irregular thinning with	Irregular thinning and
		cellular	cellular	modest reductions in cell	conspicuous reductions in
				density	cell density
1					

Figure 8: Example with multi-line text.

	Size	Grade	Involved Iymph Rode	BR status	MER-2 status	Ki-67 sbatus	LVI	MVD	Loco-regional treatment	Systemic treatment
	(fevalue)	(#walue)	(#walue)	(Rwalue)	(Mwalue)	(Analue)	(Revalue)	(Awalue)	(Revalue)	(Asvalue)
ge (30 />50 years)	3058	8236	6495	8297	34054	6641	36/13	6238	84261	<0.001
e (#1 44/21-50/>50 mm)		(84) 33	<0.001	89 79	6626	6£ 41	e0.001	0.668	«0.001	9#38
sde (1/11/11)			9258	0453	40.001	<0.001	40.001	9.0 72	80 let)	6627
olved lymph de (0/1-3/>3)				018 57	9.623	0035	«0 .001	97607	0:001	8013
gesterone (perceptor) (us (PR-/PR+)					8.002	0\$40	3692	9997	31256	80804
R-2 -/HER-2+)						0.020	0.0019	0.627	00863	34603
57 proliferative activity w Ki-67/High Ki-67)							0.049	0.041	0.698	8.089
(Absent/Present)								0.0015	3.831	34655
D (tertiles 1, 2, 3)									3:028	6.402
o-regional treatment mpectomy + radiotherapy/ stectomy + radiotherapy/	I									0.067
bleFormer predicted b	ounding b	oxes		-			. (7)			-
	1 Size	Grade	Involved lymph node	status	HER-2 status	* Ki-67 status	LVI	MVD	Loco-regional treatment	Systemic treatment
	(P-value)	(P-value)	(P-value)	(P-value)	(P-value)	1(P-value)	¤(P-value)	P-value)	(Pvalue)	(P-value)
(≤50/>50 years)	\$ 058	8 £36	#9 .495	0.296	28.054	\$ 641	8 .913	01238	£ 261	⊲0.001
≤20/21-50/>50 mm)	34	8033	≈0.001	0.879	10 .826	9 9041	40.001	81568	4 0 .001	0#438
e (1/11/11)	65	46	4 .258	0.763	≪0.001	#0.001	s:0.001	8:072	8088	3627
ed lymph		17	58	0.1597	9.323	0.035	©0.001	8.5 07	6:001	3.013
esterone -receptor	(19	70	0.002	8 :340	₹.392	0.997	0.256	0.804
-2 status	(79	0	81 82		0 .020	8 4 119	6 527	0 €63	0.803
7 prolimitative activity	50		91	(au		94	99049	0.041	875.98	01289
Absent/Present	1	01 102	105	104		95		0.015	(MART)	0.0651
(tertile	112	113	214		Tals	116][13	7][110	20628	0.602
>registed treatment pectomy + radiotherapy tectomy + radiotherapy)	183	124	125		26	127 1	20 3	29	130	0,8 57
ost-processed boundir	g boxes									
	size	Grade	nvolved ymph	status	HER-2 status	di-67 status	ΩV.	MVD	teco-regional treatment	systemic meatment
	(Mewalue)	(Analue)	node Menalue)	(Awalue)	(Pivalue)	(Mavalue)	(Levalue)	jiwalue)	walue)	(Awalue)
e (Sever)>50 years	36058	89236	59495	82 97	SR /54	90541	96/12	\$92.3 8	84261	e0.001
e 📢 🖗 / 21-50 / 550 mm		80 33	<0.001	88679	98 326	\$30 41	40.001	5000	40.001	5684
de (1/11/11)			8258	9 #/53	40.001	#0.001	40.001	\$\$ /2	SO RM	8627

Low Cut-Off Frequency
High Cut-Off Frequency
Input-Referred Noise
emrr
Number of Analog Channels
Power Consumption
Precision

PDF Cells

TableFormer predicted bounding boxes

Parameter

Gain

Parameter	value
Gain	2851 V/V (69.09 dB)
Low Cut-Off Frequency	285 Hz
High Cut-Off Frequency	6580 Hz
Input-Referred Noise	$2.1 \mu V (rms)$
EMRR	1 🗗 dB @ 1 KHz
Number of Analog Channels	2
Power Consumption	1 m A @ 3.0 V (3 mW)
Precision	Selectable, 12 or 8 bits

Value

2851 V/V (69.09 dB)

285 Hz 6580 Hz 2.1 ptV (rms) 10 dB @ 1 KHz 29 14 mA @ 3.0 V (3 mW) Selectable, 12 or 8 bits

Post-processed bounding boxes

Parameter	Value
Gain	2851 V/V (69.09 dB)
Low Cut-Off Frequency	285 Hz
High Cut-Off Frequency	6580 Hz
Paput-Referred Noise	2.1 (PV (rms)
EMRR	1ª10 dB (a) 1 KHz
Number of Analog Channels	2
Power Consumption	#mA @ 3.0 V (3 mW)
Precision	Selectable, 12 or 8 bits

TableFormer predicted structure

0 (0,	j 1 (1, 0
Parameter	Value
2 [0,] 3 (1, 1
Gain	2851 V/V (69.09 dB)
4 (0.)	5 [1, 2
Low Cut-Off Frequency	285 Hz
6 (0,	ı) 7* (1, 3
High Cut-Off Frequency	
8* (0, -	9 [1, 4
	6580 Hz
10 (0,	i] 20 [1, 5
Input-Referred Noise	2.1µV (rms)
12 (0,	i] 13 (1, 6
CMRR	110 dB @ 1 KHz
14 (0,1	rj 15 (1, 7
Number of Analog Channels	2
16 (0,	ı] 17 (1, 8
Power Consumption	1mA @ 3.0 V (3 mW)
18 (0.1	l) 19 (1, 9
Precision	Selectable, 12 or 8 bits

		1 100101011	concottable,	12 01 0 0100
ole.	Figure 10:	Example of how po	st-proce	ssing h

Figure 10: Example of how post-processing helps to restore
mis-aligned bounding boxes prediction artifact.

6° K 6	Size	Grade	Involvedlymphnode	PRstatus	HER-2status	Ki-67status	LVI	MVD	Loco- regionaltreatment	Systemictreatm
TP 8.3	(P-value)	(P-value)	(P-value)	(P-value)	P-value)	(P-value)	(P-value)	(P-value)	(P-value)	(P-value)
22 p. s	23 p	24 (2.3	25 p. s	26 jų 2	27 (6.2	26 y, z	29 7.3	30 p. :	31 (6.2	
ge (<50/>50 years)	0.058	0.236	0.495	0.297 37 K S	38(5.)	0.641 39 K S	0.913	416.1	42 14 1	<0.001
lize (≤20/21-50/>50 nm)		0.033	<0.001	0.879	0.826	0.041	<0.001	0.568	<0.001	0.438
44 p. 4	45*0	4 46 2.4	47 p. q	44 K 4	4915.4	50 x 4	\$17.4	0.070 ⁽² A)	5 A 0 0	0.007
srade (VIVIII)	99" (1	s 57° µ. c	0.236	0.753 NK	0.001	50.001 61 jk s	42.001	0.072 63 jul	0.100	0.627
nvolved lymphnode 0/1-3/>3)				0.157	0.323	0.035	<0.001	0.607	0.001	0.013
rogesterone- eceptorstatus (PR- PR+)	67° p	.e 68'9.4	662 p. 4	20° ji k	0.002	0.340	0.392	0.997	0.256	0.804
IER-2 status(HER-2 - HER-2+)	78"(1	n 114 µ.1	80° (t. 1)	61° K 3	62" (5.7	0.020	0.119	0.627	0.363	0.603
i-67 proliferative ctivity(Low Ki-67/High i-67)	60° p	n 60° p. s	91° p. q	92° y. s	92° p. s	94° y. s	0.049	90 p. 1 0.041	0.598	0.289
VI (Absent/Present)	1 900° (r	a 101° p. s	102" [3.9]	100° ji n	904° (s. s	106° ji 9	106° (r. sj	0.215	0.031	0.255
190 p. vo IVD (tertiles 1, 2, 3)	100.00	112° p. 1	113° p. 10	154° (x. 18	115° p. u	116° p. 10	117° (r. 11	1187 (6.11	0.628	0.402
oco-regional eatment(Lumpectomy radiotherapy/	122"().	ni 123° p. n	724" (5. 9)	128° (K. H)	128° (5. m	127° (K. H	128° p. H	129" (8. 1	1 1307 (K. H	0.057

51002 51840 91692 51897

1402U 140319 140427

349419 54941

JAN 1

Hogesterone Caseptor Ratus (PR-/PR+)

1067 proliferative activity 109w Ki-67/High Ki-67)

(Absent/Present)

101R-2 -/HER-2+)

\$8256

4.953

52598

0.0031 0.0028 9.604

\$1603

31265

30803 30802

PDF Cells				TableFormer predicted bounding boxes				Post-processed bounding boxes				TableFormer predicted structure			
tem		Farmers	General	ltem		Farmers	² General community	tem		Farmers	² General community	ltem	0 ; 3* ;	Farmers	Generalcommunity
Socio-economic			Socioneconomic		4		Socio-e	conomic			6p.	7:	8		
åge (96)	8-29 yrs	9.20	2.55	Age (96)	18-29 yrs	8.20	2.55	age (%)	8-29 yrs	1 .20	2.55	Age (%)	18-29 yrs 11 r	8.20	12.55
	30-49 yrs	47.80	38.41	10	49 VI6	#7.80	38341		30 -49 yrs	19 ,80	38,41	141 m	30-49 yrs	47.80	38.41
	33+ yrs	44.00	49.05	14	SH+ VIS	44.00	49.05		58+ yrs	44.00	49:05			44.00	49.05
30x (%)	Male	12.80	49.96	38× (96)	Male	20	21	M X (96)	Male	12.80	48.96	Sex (%)	50+ yrs	72.80	s 21° 0.6
	Bemale	20.20	50.04		Eemale.	22-20	14404		aemale	£¥.20	20.04	22* p.	Male 23 r	e 24° p.	49.96
acation (%)	Major metropolitan	24	28.35	Proceeding (96)	27 Maior metropolitan		58.35	acation (%)	Major metropolitar	3224	36.35	26 n.	EamaleMajor metropolita	28 g.	29 p. 1
	Memainder of State	100.00	46.65		Remainder of State	100-00	31 65		semainder of State	280.00	44.65	30° p.	e 31 ;	e 32 g.	а 33 р.е
Stationship Status (%)	Married	35.80	56.19	Relationship	Married	¥6 .80	98 .19	stelationship steltus (%)	Married	5 6. 80	56.19	34 p. Relationshipstatu	Married	75.80	41.65 8 37 p. 9 56.19
	mgle	34.20	46.81		Sangle	29.20	43.81		angle	# 1.20	46.81	(76) 38° p. 1	a 30 p.	ag 40 p. :	41 p. 10
work status (%)	Hall-time	30.60	40.23	Work status (%)	Full-time	80.60	49.23	Alork status (%)	Mail-time	99.60	Pat.25	42.0.1	Single 43 K	24.20	43.81 •• 45 p. +•
	Part-time	19.40	99.12		(Part-time)	16.40	19.12		Part-time	19.40	19.12	Work status (%)	Full-time	80.60	42.23
	Not in paid work	45	38.65	50	Not in paid work	SI.	865		Not in paid work	Bas .	38.65	-	Part-time	19.40	19.12
evel of	Begree	14.20	31.75	€€vel of education.(%)	Segree	14.20	5? .75	Movel of	Segree	14.20	53.75	54 p.1	Not in paid work	n 62 (2.1	38.65
	S iploma	38.80	54.83	58	59 Diploma	50 38.80	19 83		Mploma	38.80	88.85	(%)	Degree	14.20	31.75
	¥ear 11	30.60	\$8.38	62	Pear 11	09.60	16.38		Mear 11	18 .60	¥6.38	68° ju i	Diploma	38.80	33.83 61 p. 16
	¥ear 10	¥7.20	\$0.31	66	Mar 10	FF.20	1031		Mear 10	17.2 0	14.31	62° p. s	e 63 p. Year 11	20.60	16.38
	Vear 9 or below	9220	65'3		Year 9 or below	73.20	R 73		Wear 9 or below	\$220	89/5	66° ju i	67 s.	17.00	a 69 p. 11
eusehold income	4 \$30,000	¥7.40	34 .03	Nousehold income	<\$30,000	75 .40	299.03	Constant Con	48 \$30,000	67.40	14.0 2	70° p. 1	Year 9 or below	9.20	6.73
	\$30,000-\$79,999	58.20	42.81	78	73 0.000-\$79.999	50.20	42.81		330,000 37 49,995	10. 20	4008	74 js. 1 Household	< \$30,000	17.40	21.03
	380,000+	39.40	36.16		\$80,000+	32.40	36.16		98 0,000+	39.40	36.16	income(%) 78' p. 2	o 79 p.	an 80 p. :	119 p. 20
Satisfaction			Satisfaction				Satisfaction				\$30,000-\$79,999 50.20 42.81				
24 erall (M, SD) 24.55 (#6.64)			everal (M, SD)	84	4955° (17.88) 77.25° (16.64)		werall (M, SD)		14.55 (47.83 (7.25 (86.64)		Satisfaction				
												Overall (M, SD)	a 447).	74.55b(17.83)	77.25(16.64)
Թ main (M, SD)	<pre>immunity monnectedness</pre>	33 .87 ⁹ [886.64)	66.57 (18.92)	Bomain (M, SD)	Connectedness	9 .87 ^a (18.64)	69.57 (19.92)	🕬 main (M, SD)	sommunity sonnectedness	33 .87 166 .64	99 .57 (199.92)	87 p. 2 Domain (M, SD) 91" p. 2	Communityconnectedne:	89 (c. 1 15 73.87a(18.64)	69.57(19.92)
	Welationships	83.17 ⁹ (%7.63)	39 .49 (32 .79)	31	Relationships	83.17 ^a (17.63)	7949 (22.79)		Relationships	33.17 10 7.63	19 .49 (32.79)	961	Relationships	83.17a(17.63)	79.49(22.79)
	sefety	33.61 *(803.54)	X82 94 (10 84)		Safety	83.61 ^a (16.54)	78994 (17.84)		safety	33 .61 (86.54)	ANSIA (100 84)		Safety	83.61a(16.54)	78.94(17.84)
	Stendard of living	14946 1(107.91)	77466 ⁴ (940.65)	99	Mandard of living	74746 ⁰ (16.91)	17,66° (16,65)		stendard of living	14446 ¹ 103.91	7466 ⁴ [740.65]	99° N 2	Standard of living	74.46b(16.91)66.8	.9 77.66a(16.65)
	Bature security	06839 ³⁴ (08.34)	X4539 (30612)	103	Future security	85 .89 ⁶ (21.34)	74939 (20.12)		Henure security	10859 ¹⁴ 10934	X4439 (JA0412)	103° ji z	Future security	b(21.34)76.80	71.39(20.12)
	Maraith	166 30 1 669.67)	34133 (19222)	111	kiealth	10 80° (17.67)	74.33 (19.22)		MARAIT	10030 1000.67	1413 \$ (1 188 22)	107° je s	s 100 p.	× 120 p.:	121 p.m.
	Athieving in life	12:56 (18:54)	7409ª (88.54)		Achieving in life	72.36 (18.34)	74.09 ^a (18.54)		Athieving in life	12836 (198934)	1409 188.54	111° p. z	112 p.	an 113 p. 1	n 114 p.m
	Religion/spirituality	350 34 3503 .24)	08023 ¹¹ 996.21)		Religion/spirituality	65.34 ^b (26.24)	68.23 ^b (26.21)		Religion/spirituality	338 34 1398 .24	3823 4 86.21	115" (c. 5	Achieving in life	72.36(18.34)	74.09a(18.54)
													Religion/spirituality	65.34b(26.24)	68.23b(26.21)

Figure 11: Example of long table. End-to-end example from initial PDF cells to prediction of bounding boxes, post processing and prediction of structure.