

Supplementary: Treading Towards Privacy-Preserving Table Structure Recognition

Sachin Raja
IIIT Hyderabad

sachin.raja@research.iiit.ac.in

Ajoy Mondal
IIIT Hyderabad

ajoy.mondal@iiit.ac.in

C V Jawahar
IIIT Hyderabad

jawahar@iiit.ac.in

1. Dataset Preprocessing

Our model for cell detection is trained on the FinTab-Net dataset, which initially provides bounding boxes around cell content. To transform this dataset into one with cell-level coordinates, we implement a preprocessing step. This involves determining the minimum and maximum x and y coordinates for each row and column based on the content boxes. Subsequently, we refine the coordinates by splitting the y-coordinate difference between adjacent rows by half. This extends the maximum y end coordinate of the preceding row while reducing the minimum y-coordinate of the current row. Similarly, we split the x-coordinate difference between adjacent columns by half, extending the maximum x end coordinate of the previous column and reducing the minimum x-coordinate of the current column. This processing ensures that our dataset adheres to all constraints relevant to our modeling approach. Figure 1 visualizes the steps.

2. Postprocessing Cell Detection Output

Following the identification of table cell bounding boxes and the prediction of row/column adjacency matrices, the subsequent task is to establish row and column spans for each cell, facilitating the reconstruction of the table’s structure. Our post-processing algorithm, an extension of previously proposed methodologies [1, 2], integrates word/token bounding boxes extracted through OCR tools to refine cell boxes and enhance structural recognition. The algorithm entails several steps, including the removal of table cells with heights below 60 percent of the average height of predicted cells, the identification and exclusion of incorrectly predicted cells based on various loss weight values and overlap considerations, and the definition of functions for calculating overlap and containment between cell pairs. Additionally, the algorithm involves sorting all table cells based on their x-end coordinates and iterating through them to assign column indices, considering overlapping cells and applying threshold criteria. Ultimately, the resulting columns list for each cell aids in determining the start and

	Options (in thousands)	Wtd. Avg. Exercise Price	Wtd. Avg. Remaining Contractual Term (in years)	Aggregate Intrinsic Value (dollars in millions)
Outstanding at April 25, 2014	35,577	\$ 44.78		
Granted	40,952	57.96		
Exercised	(13,503)	45.32		
Expired/Forfeited	(1,005)	50.43		
Outstanding at April 24, 2015	62,021	53.27	7.03	\$ 1,351
Vested and expected to vest at April 24, 2015	55,649	51.27	6.75	1,314
Exercisable at April 24, 2015	28,272	39.91	4.53	981

(a) Original Table Image

	Options (in thousands)	Wtd. Avg. Exercise Price	Wtd. Avg. Remaining Contractual Term (in years)	Aggregate Intrinsic Value (dollars in millions)
Outstanding at April 25, 2014	35,577	\$ 44.78		
Granted	40,952	57.96		
Exercised	(13,503)	45.32		
Expired/Forfeited	(1,005)	50.43		
Outstanding at April 24, 2015	62,021	53.27	7.03	\$ 1,351
Vested and expected to vest at April 24, 2015	55,649	51.27	6.75	1,314
Exercisable at April 24, 2015	28,272	39.91	4.53	981

(b) Original Cell Annotations

	Options (in thousands)	Wtd. Avg. Exercise Price	Wtd. Avg. Remaining Contractual Term (in years)	Aggregate Intrinsic Value (dollars in millions)
Outstanding at April 25, 2014	35,577	\$ 44.78		
Granted	40,952	57.96		
Exercised	(13,503)	45.32		
Expired/Forfeited	(1,005)	50.43		
Outstanding at April 24, 2015	62,021	53.27	7.03	\$ 1,351
Vested and expected to vest at April 24, 2015	55,649	51.27	6.75	1,314
Exercisable at April 24, 2015	28,272	39.91	4.53	981

(c) Intermediate Annotations

	Options (in thousands)	Wtd. Avg. Exercise Price	Wtd. Avg. Remaining Contractual Term (in years)	Aggregate Intrinsic Value (dollars in millions)
Outstanding at April 25, 2014	35,577	\$ 44.78		
Granted	40,952	57.96		
Exercised	(13,503)	45.32		
Expired/Forfeited	(1,005)	50.43		
Outstanding at April 24, 2015	62,021	53.27	7.03	\$ 1,351
Vested and expected to vest at April 24, 2015	55,649	51.27	6.75	1,314
Exercisable at April 24, 2015	28,272	39.91	4.53	981

(d) Final Preprocessed Annotations

Figure 1. Visualization of the Annotation Preprocessing steps to ensure alignment and continuity

end column indices, contributing to the comprehensive reconstruction of the table’s layout.

Similarly, two functions are defined to calculate overlap and containment for cell pairs along the vertical direction, using ratios of intersecting line segment lengths. The sub-

sequent steps involve sorting all table cells based on their y-end coordinates and addressing the challenge of multi-row spanning cells by initializing empty lists for each cell and setting the current row index to zero. The iterative process involves identifying overlapping cells for each cell, updating the rows list based on specific criteria, and incrementing the row index accordingly. The resultant rows list facilitates the determination of start and end row indices for every cell.

Moving on to the smoothing and fine-tuning phase, row and column indices, along with bounding boxes, are utilized to align start and end coordinates. For columns, the process involves extracting minimum start-x coordinates and maximum end-x coordinates, dividing the gap equally among adjacent cells to ensure adherence to continuity, alignment, and non-overlapping constraints. The algorithm further leverages an OCR tool, such as Tesseract [3], for pre-processing the table image and obtaining word/token-level bounding boxes. The final step involves refining the predicted cell bounding boxes using token bounding boxes, computing overlapping areas, assigning them to the highest overlapping cell, and adjusting cell boundaries along both x and y directions to fully encapsulate the content. This meticulous approach notably enhances detection performance, particularly at higher Intersection over Union (IoU) threshold values.

3. Optimization Behavior

Figures 2, 3, 4, 5 and 6 compare optimization behaviors of using Table Grid Approximator (TGA) vs Region Proposal Network (RPN) on classification, bounding-box regression, alignment, row-continuity and column-continuity losses respectively. Consistently across all five loss types, it is evident that using TGA allows for optimization to a lower loss value in a lesser time as compared to using RPN.

4. Distribution of Anchor Boxes

As opposed to a fix number of anchors set as 20,000, using TGA allows dynamically generated anchors which depend on the approximated grid size of the image. Figure 7 shows the probability distribution of text contours, grid cells and anchors respectively on linear as well as log scale. The graphs show that TGA allows for significantly reduced number of anchors on average as opposed to a fixed number of 20,000.

5. Annotation Errors in FinTabNet Train

Figures 8 through 12 demonstrate some erroneously annotated images in the FinTabNet-Train dataset.

6. Qualitative Examples

Figures 13 through 24 demonstrate near-perfect structure prediction on dense tables as well as those containing empty cells.

6.1. Success Cases

6.2. Failure Cases

Figures 25 through 28 demonstrate a few failure cases, which despite being predicted incorrectly are not very far from the ground truth.

6.3. Errors in Annotations in FinTabNet-Test

Figures 29 through 31 demonstrate some examples on FinTabNet-Test dataset, which despite having incorrect annotations, were predicted accurately by our solution. This further demonstrates the robustness of TabGuard.

References

- [1] Sachin Raja, Ajoy Mondal, and CV Jawahar. Visual understanding of complex table structures from document images. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 2299–2308, 2022. 1
- [2] Sachin Raja, Ajoy Mondal, and C. V. Jawahar. Table structure recognition using top-down and bottom-up cues. In *European Conference on Computer Vision (ECCV)*, pages 70–86, 2020. 1
- [3] Ray Smith. An overview of the Tesseract OCR engine. In *ICDAR*, 2007. 2
- [4] Xinyi Zheng, Douglas Burdick, Lucian Popa, Xu Zhong, and Nancy Xin Ru Wang. Global Table Extractor (GTE): A framework for joint table identification and cell structure recognition using visual context. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 697–706, 2021. 5, 6, 7, 8, 9, 10, 11, 12, 13, 14

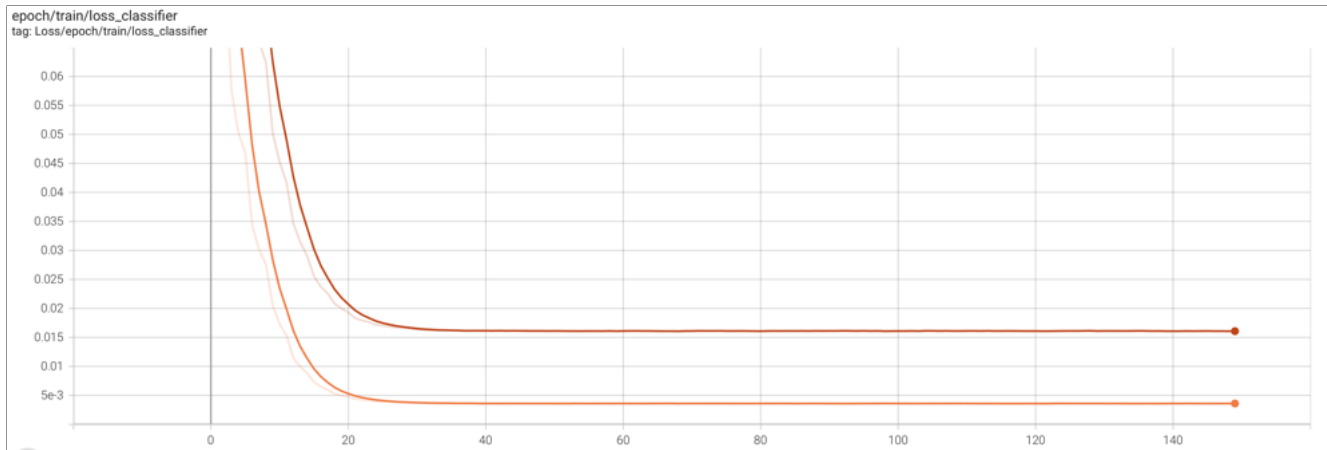


Figure 2. Optimization behavior for the classification loss. Orange curve represents loss using Table Grid Approximator (TGA) instead of Region Proposal Network (RPN), while the red curve represents loss using RPN instead of TGA.

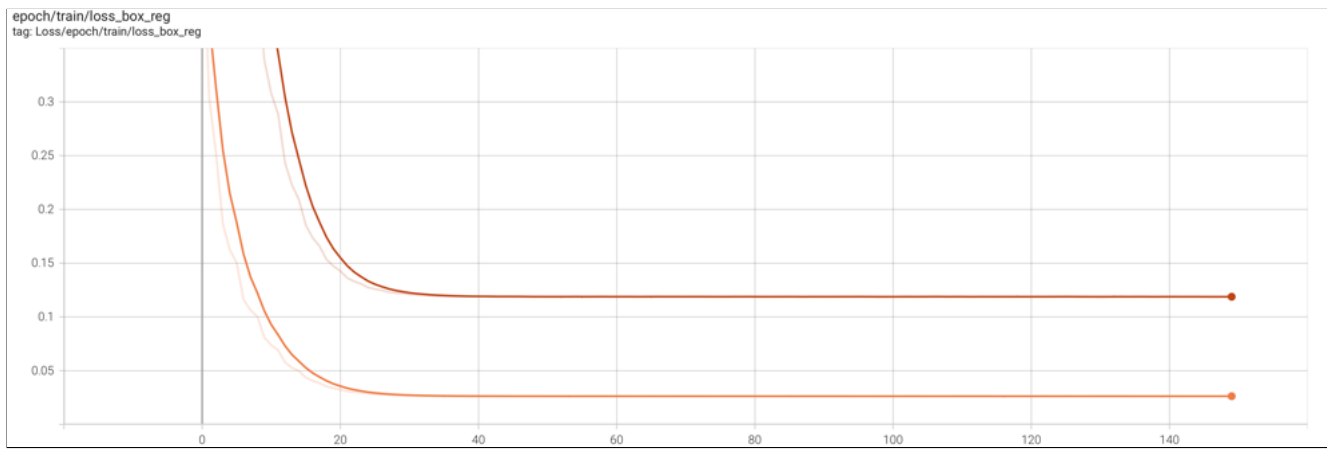


Figure 3. Optimization behavior for the bounding box regression loss. Orange curve represents loss using Table Grid Approximator (TGA) instead of Region Proposal Network (RPN), while the red curve represents loss using RPN instead of TGA.

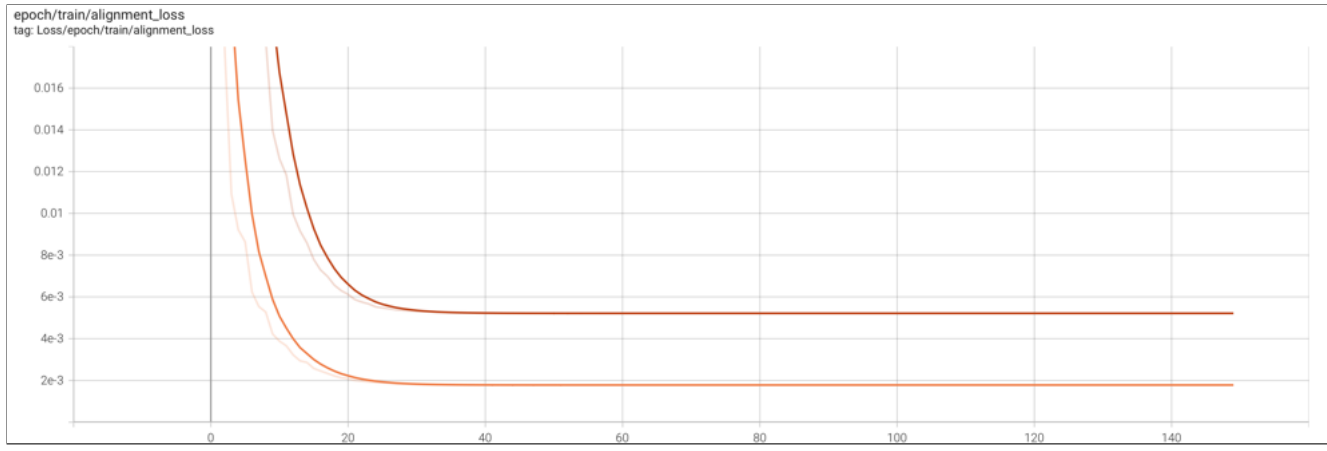


Figure 4. Optimization behavior for the alignment loss. Orange curve represents loss using Table Grid Approximator (TGA) instead of Region Proposal Network (RPN), while the red curve represents loss using RPN instead of TGA.

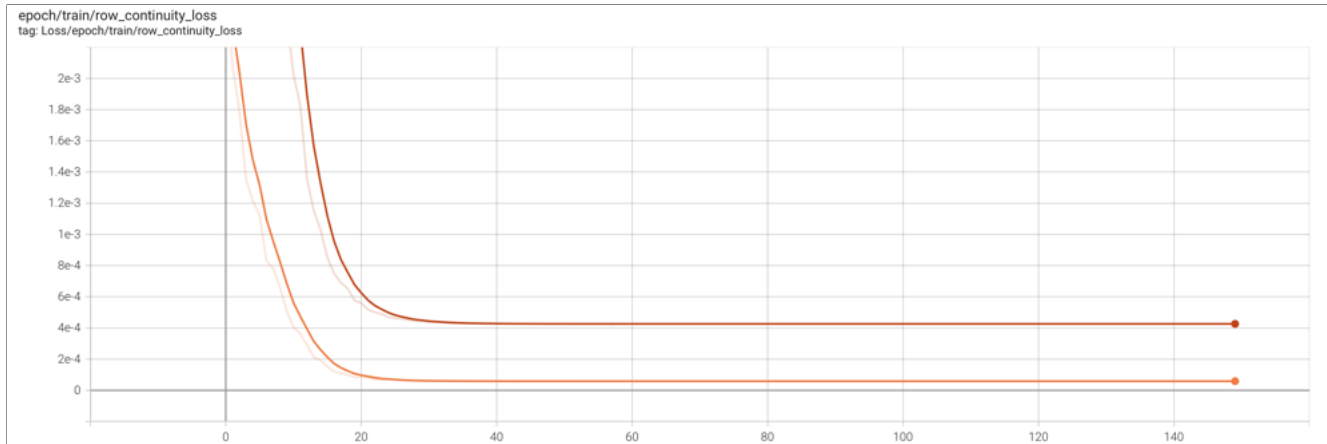


Figure 5. Optimization behavior for the row continuity loss. Orange curve represents loss using Table Grid Approximator (TGA) instead of Region Proposal Network (RPN), while the red curve represents loss using RPN instead of TGA.

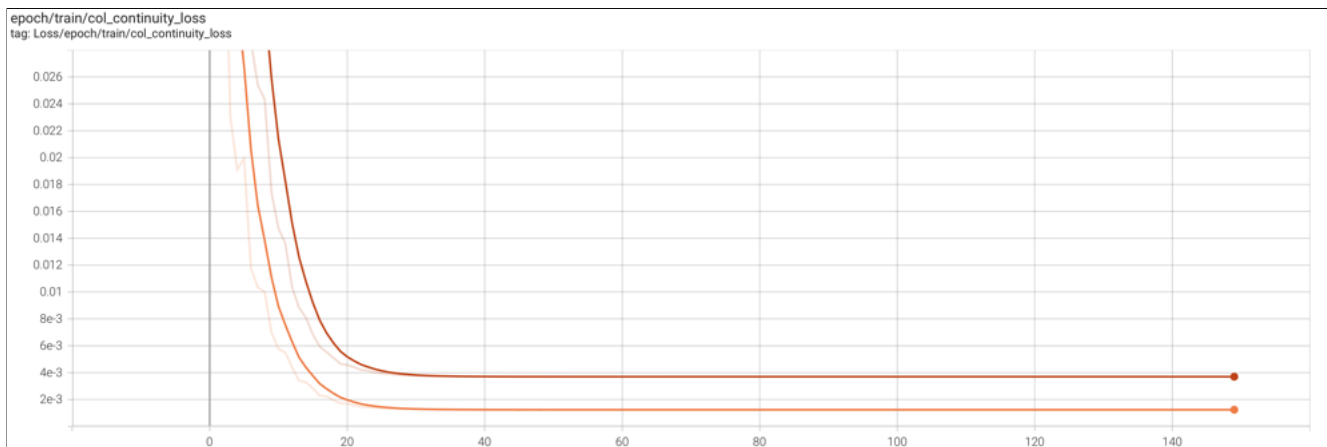


Figure 6. Optimization behavior for the column continuity loss. Orange curve represents loss using Table Grid Approximator (TGA) instead of Region Proposal Network (RPN), while the red curve represents loss using RPN instead of TGA.

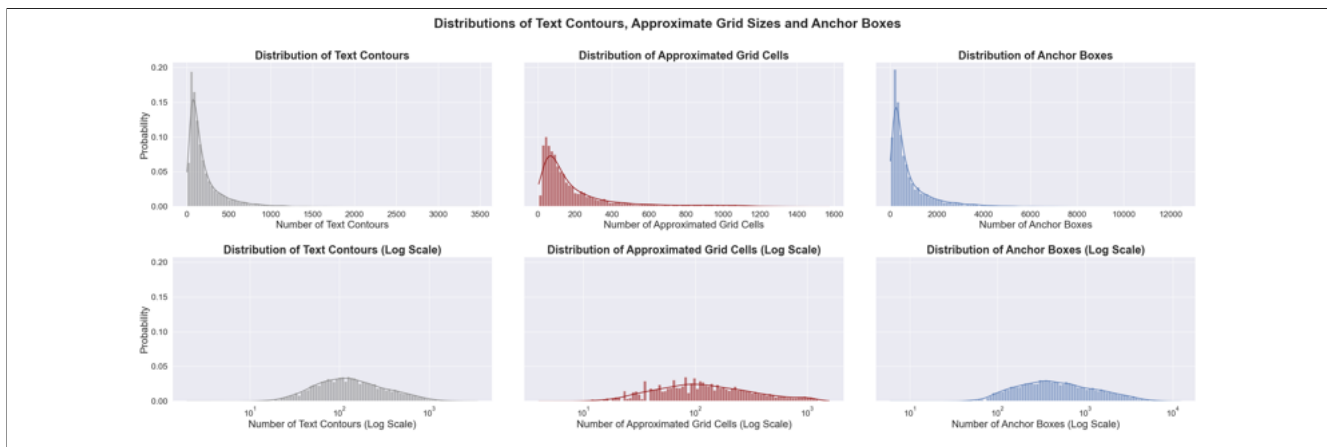


Figure 7. Distributions of text contours, table grid cells and generated anchors. X-axes denote the counts of contours, cells, and anchors respectively, while Y-axis denotes probability of occurrence. Row 1 is on linear scale while Row 2 is on the log scale. It is clear that text contours, grid cells and anchor boxes all follow log-linear distributions.

Figure 8. An example of file from Fintabnet-train [4] dataset which is incorrectly annotated. Reading order of images proceeds from left to right and then to the next row. First image shows the original image, second image shows the original annotation, third image shows the intermediate step of annotation preprocessing, and fourth image shows the preprocessed annotation.

Figure 9. An example of file from Fintabnet-train [4] dataset which is incorrectly annotated. Reading order of images proceeds from left to right and then to the next row. First image shows the original image, second image shows the original annotation, third image shows the intermediate step of annotation preprocessing, and fourth image shows the preprocessed annotation.

(1) Financial Statements	Page No.	(1) Financial Statements	Page No.
(a) Report of Independent Registered Public Accounting Firm	77	(a) Report of Independent Registered Public Accounting Firm	77
(b) Consolidated Statements of Income for the years ended December 31, 2014, 2013 and 2012	78	(b) Consolidated Statements of Income for the years ended December 31, 2014, 2013 and 2012	78
(c) Consolidated Statements of Comprehensive Income for the years ended December 31, 2014, 2013 and 2012	79	(c) Consolidated Statements of Comprehensive Income for the years ended December 31, 2014, 2013 and 2012	79
(d) Consolidated Balance Sheets as of December 31, 2014 and 2013	80-81	(d) Consolidated Balance Sheets as of December 31, 2014 and 2013	80-81
(e) Consolidated Statements of Cash Flows for the years ended December 31, 2014, 2013 and 2012	83	(e) Consolidated Statements of Cash Flows for the years ended December 31, 2014, 2013 and 2012	83
(f) Consolidated Statements of Shareholder's Equity for the years ended December 31, 2014, 2013 and 2012	84-85	(f) Consolidated Statements of Shareholder's Equity for the years ended December 31, 2014, 2013 and 2012	84-85
(g) Notes to Consolidated Financial Statements	86-135	(g) Notes to Consolidated Financial Statements	86-135
(2) Financial Statements Schedules		(2) Financial Statements Schedules	
All schedules have been omitted because of the absence of conditions under which they are required.		All schedules have been omitted because of the absence of conditions under which they are required.	

(1) Financial Statements	Page No.	(1) Financial Statements	Page No.
(a) Report of Independent Registered Public Accounting Firm	77	(a) Report of Independent Registered Public Accounting Firm	77
(b) Consolidated Statements of Income for the years ended December 31, 2014, 2013 and 2012	78	(b) Consolidated Statements of Income for the years ended December 31, 2014, 2013 and 2012	78
(c) Consolidated Statements of Comprehensive Income for the years ended December 31, 2014, 2013 and 2012	79	(c) Consolidated Statements of Comprehensive Income for the years ended December 31, 2014, 2013 and 2012	79
(d) Consolidated Balance Sheets as of December 31, 2014 and 2013	80-81	(d) Consolidated Balance Sheets as of December 31, 2014 and 2013	80-81
(e) Consolidated Statements of Cash Flows for the years ended December 31, 2014, 2013 and 2012	83	(e) Consolidated Statements of Cash Flows for the years ended December 31, 2014, 2013 and 2012	83
(f) Consolidated Statements of Shareholder's Equity for the years ended December 31, 2014, 2013 and 2012	84-85	(f) Consolidated Statements of Shareholder's Equity for the years ended December 31, 2014, 2013 and 2012	84-85
(g) Notes to Consolidated Financial Statements	86-135	(g) Notes to Consolidated Financial Statements	86-135
(2) Financial Statements Schedules		(2) Financial Statements Schedules	
All schedules have been omitted because of the absence of conditions under which they are required.		All schedules have been omitted because of the absence of conditions under which they are required.	

Figure 10. An example of file from Fintabnet-train [4] dataset which is incorrectly annotated. Reading order of images proceeds from left to right and then to the next row. First image shows the original image, second image shows the original annotation, third image shows the intermediate step of annotation preprocessing, and fourth image shows the preprocessed annotation.

(1) Financial Statements	Page No.	(1) Financial Statements	Page No.
(a) Reports of Independent Registered Public Accounting Firms	67-68	(a) Reports of Independent Registered Public Accounting Firms	67-68
(b) Consolidated Statements of Income for the years ended December 31, 2008, 2007 and 2006	69	(b) Consolidated Statements of Income for the years ended December 31, 2008, 2007 and 2006	69
(c) Consolidated Balance Sheets as of December 31, 2008 and 2007	70-71	(c) Consolidated Balance Sheets as of December 31, 2008 and 2007	70-71
(d) Consolidated Statements of Cash Flows for the years ended December 31, 2008, 2007 and 2006	73	(d) Consolidated Statements of Cash Flows for the years ended December 31, 2008, 2007 and 2006	73
(e) Consolidated Statements of Shareholders' Equity and Comprehensive Income for the years ended December 31, 2008, 2007 and 2006	74-75	(e) Consolidated Statements of Shareholders' Equity and Comprehensive Income for the years ended December 31, 2008, 2007 and 2006	74-75
(f) Notes to Consolidated Financial Statements	76-117	(f) Notes to Consolidated Financial Statements	76-117

(1) Financial Statements	Page No.	(1) Financial Statements	Page No.
(a) Reports of Independent Registered Public Accounting Firms	67-68	(a) Reports of Independent Registered Public Accounting Firms	67-68
(b) Consolidated Statements of Income for the years ended December 31, 2008, 2007 and 2006	69	(b) Consolidated Statements of Income for the years ended December 31, 2008, 2007 and 2006	69
(c) Consolidated Balance Sheets as of December 31, 2008 and 2007	70-71	(c) Consolidated Balance Sheets as of December 31, 2008 and 2007	70-71
(d) Consolidated Statements of Cash Flows for the years ended December 31, 2008, 2007 and 2006	73	(d) Consolidated Statements of Cash Flows for the years ended December 31, 2008, 2007 and 2006	73
(e) Consolidated Statements of Shareholders' Equity and Comprehensive Income for the years ended December 31, 2008, 2007 and 2006	74-75	(e) Consolidated Statements of Shareholders' Equity and Comprehensive Income for the years ended December 31, 2008, 2007 and 2006	74-75
(f) Notes to Consolidated Financial Statements	76-117	(f) Notes to Consolidated Financial Statements	76-117

Figure 11. An example of file from Fintabnet-train [4] dataset which is incorrectly annotated. Reading order of images proceeds from left to right and then to the next row. First image shows the original image, second image shows the original annotation, third image shows the intermediate step of annotation preprocessing, and fourth image shows the preprocessed annotation.

Notes Issued (in millions)	Maturity Date	Interest Rate	Interest Payment Dates
\$ 750	May 18, 2016	Variable; three-month U.S. dollar LIBOR, reset quarterly, plus 45 basis points	February 18, May 18, August 18, and November 18; commencing on February 18, 2015
750	November 17, 2017	Fixed 1.750%	May 17 and November 17; commencing on May 17, 2015
1,250	November 18, 2019	Fixed 2.700%	May 18 and November 18; commencing on May 18, 2015
1,250	November 18, 2021	Fixed 3.300%	May 18 and November 18; commencing on May 18, 2015
2,000	November 18, 2024	Fixed 3.800%	May 18 and November 18; commencing on May 18, 2015
500	November 18, 2034	Fixed 4.500%	May 18 and November 18; commencing on May 18, 2015
1,500	November 18, 2044	Fixed 4.800%	May 18 and November 18; commencing on May 18, 2015
\$8,000			

Notes Issued (in millions)	Maturity Date	Interest Rate	Interest Payment Dates
\$ 750	May 18, 2016	Variable; three-month U.S. dollar LIBOR, reset quarterly, plus 45 basis points	February 18, May 18, August 18, and November 18; commencing on February 18, 2015
750	November 17, 2017	Fixed 1.750%	May 17 and November 17; commencing on May 17, 2015
1,250	November 18, 2019	Fixed 2.700%	May 18 and November 18; commencing on May 18, 2015
1,250	November 18, 2021	Fixed 3.300%	May 18 and November 18; commencing on May 18, 2015
2,000	November 18, 2024	Fixed 3.800%	May 18 and November 18; commencing on May 18, 2015
500	November 18, 2034	Fixed 4.500%	May 18 and November 18; commencing on May 18, 2015
1,500	November 18, 2044	Fixed 4.800%	May 18 and November 18; commencing on May 18, 2015
\$8,000			

Figure 12. An example of file from Fintabnet-train [4] dataset which is incorrectly annotated. Reading order of images proceeds from left to right and then to the next row. First image shows the original image, second image shows the original annotation, third image shows the intermediate step of annotation preprocessing, and fourth image shows the preprocessed annotation.

Millions of dollars and shares except per share data	Year Ended December 31		
	2008	2007	2006
Revenue:			
Services	\$ 13,391	\$ 11,256	\$ 9,643
Product sales	4,888	4,008	3,312
Total revenue	18,279	15,264	12,955
Operating costs and expenses:			
Cost of services	10,079	8,167	6,751
Cost of sales	3,970	3,358	2,873
General and administrative	282	293	342
Gain on sale of business assets, net	(82)	(52)	(78)
Total operating costs and expenses	14,269	11,766	9,788
Operating income	4,010	3,498	3,243
Interest expense	(190)	(154)	(163)
Interest income	39	124	128
Other, net	(738)	(8)	(10)
Income from continuing operations before income taxes and minority interest	3,163	3,460	3,199
Provision for income taxes	(1,213)	(907)	(1,060)
Minority interest in net income of subsidiaries	9	(29)	(18)
Income from continuing operations	1,961	2,524	2,117
Income (loss) from discontinued operations, net of income tax (provision) benefit of \$3, \$0.15, and \$(183)	(423)	975	171
Net income	\$ 1,538	\$ 3,499	\$ 2,288
Basic income (loss) per share:			
Income from continuing operations	\$ 2.24	\$ 2.76	\$ 2.13
Income (loss) from discontinued operations, net	(0.49)	1.07	0.16
Net income per share	\$ 1.75	\$ 3.83	\$ 2.31
Diluted income (loss) per share:			
Income from continuing operations	\$ 2.17	\$ 2.68	\$ 2.07
Income (loss) from discontinued operations, net	(0.47)	1.02	0.16
Net income per share	\$ 1.70	\$ 3.68	\$ 2.23
Basic weighted average common shares outstanding	877	913	1,014
Diluted weighted average common shares outstanding	904	950	1,054

Figure 13. An example of file from Fintabnet-test [4] dataset. Reading order of images proceeds from left to right and then to the next row. First image shows the original annotation, second image shows the masked image, third image shows the approximated table grid, and fourth image shows the output of TabGuard.

Figure 14. An example of file from Fintabnet-test [4] dataset. Reading order of images proceeds from left to right and then to the next row. First image shows the original annotation, second image shows the masked image, third image shows the approximated table grid, and fourth image shows the output of TabGuard.

Figure 15. An example of file from Fintabnet-test [4] dataset. Reading order of images proceeds from left to right and then to the next row. First image shows the original annotation, second image shows the masked image, third image shows the approximated table grid, and fourth image shows the output of TabGuard.

Figure 16. An example of file from Fintabnet-test [4] dataset. Reading order of images proceeds from left to right and then to the next row. First image shows the original annotation, second image shows the masked image, third image shows the approximated table grid, and fourth image shows the output of TabGuard.

This figure shows a financial statement with multiple columns and rows. The columns represent different financial metrics, and the rows represent different categories. The data is presented in a structured format, with some cells containing redacted information.

Figure 17. An example of file from Fintabnet-test [4] dataset. Reading order of images proceeds from left to right and then to the next row. First image shows the original annotation, second image shows the masked image, third image shows the approximated table grid, and fourth image shows the output of TabGuard.

This figure shows a financial statement with multiple columns and rows. The columns represent different financial metrics, and the rows represent different categories. The data is presented in a structured format, with some cells containing redacted information.

Figure 18. An example of file from Fintabnet-test [4] dataset. Reading order of images proceeds from left to right and then to the next row. First image shows the original annotation, second image shows the masked image, third image shows the approximated table grid, and fourth image shows the output of TabGuard.

This figure shows a financial statement with multiple columns and rows. The columns represent different financial metrics, and the rows represent different categories. The data is presented in a structured format, with some cells containing redacted information.

Figure 19. An example of file from Fintabnet-test [4] dataset. Reading order of images proceeds from left to right and then to the next row. First image shows the original annotation, second image shows the masked image, third image shows the approximated table grid, and fourth image shows the output of TabGuard.

Table Number	Description	Table	Description	Table	Description
3.1	Annual Certificate of Incorporation filed as an Exhibit to the Company's report on Form S-8, filed on August 18, 2011, and incorporated herein by reference. (Commission File Number 333-171112)			3.1	Annual Certificate of Incorporation filed as an Exhibit to the Company's report on Form S-8, filed on August 18, 2011, and incorporated herein by reference. (Commission File Number 333-171112)
3.2	Amended and Restated By-Laws of the Company filed as an Exhibit to the Company's report on Form S-8, filed on August 18, 2011, and incorporated herein by reference. (Commission File Number 333-171112)			3.2	Amended and Restated By-Laws of the Company filed as an Exhibit to the Company's report on Form S-8, filed on August 18, 2011, and incorporated herein by reference. (Commission File Number 333-171112)
3.3	Form of 1099 Service Note due 2011 filed as an Exhibit to the Company's report on Form S-8, filed on August 18, 2011, and incorporated herein by reference. (Commission File Number 333-171112)			3.3	Form of 1099 Service Note due 2011 filed as an Exhibit to the Company's report on Form S-8, filed on August 18, 2011, and incorporated herein by reference. (Commission File Number 333-171112)
3.4	Form of 1099 Service Note due 2012 filed as an Exhibit to the Company's report on Form S-8, filed on August 18, 2011, and incorporated herein by reference. (Commission File Number 333-171112)			3.4	Form of 1099 Service Note due 2012 filed as an Exhibit to the Company's report on Form S-8, filed on August 18, 2011, and incorporated herein by reference. (Commission File Number 333-171112)
3.5	Form of 1099 Service Note due 2013 filed as an Exhibit to the Company's report on Form S-8, filed on August 18, 2011, and incorporated herein by reference. (Commission File Number 333-171112)			3.5	Form of 1099 Service Note due 2013 filed as an Exhibit to the Company's report on Form S-8, filed on August 18, 2011, and incorporated herein by reference. (Commission File Number 333-171112)
3.6	Form of 1099 Service Note due 2014 filed as an Exhibit to the Company's report on Form S-8, filed on August 18, 2011, and incorporated herein by reference. (Commission File Number 333-171112)			3.6	Form of 1099 Service Note due 2014 filed as an Exhibit to the Company's report on Form S-8, filed on August 18, 2011, and incorporated herein by reference. (Commission File Number 333-171112)
3.7	Form of 1099 Service Note due 2015 filed as an Exhibit to the Company's report on Form S-8, filed on August 18, 2011, and incorporated herein by reference. (Commission File Number 333-171112)			3.7	Form of 1099 Service Note due 2015 filed as an Exhibit to the Company's report on Form S-8, filed on August 18, 2011, and incorporated herein by reference. (Commission File Number 333-171112)
3.8	Form of 1099 Service Note due 2016 filed as an Exhibit to the Company's report on Form S-8, filed on August 18, 2011, and incorporated herein by reference. (Commission File Number 333-171112)			3.8	Form of 1099 Service Note due 2016 filed as an Exhibit to the Company's report on Form S-8, filed on August 18, 2011, and incorporated herein by reference. (Commission File Number 333-171112)
3.9	Form of 1099 Service Note due 2017 filed as an Exhibit to the Company's report on Form S-8, filed on August 18, 2011, and incorporated herein by reference. (Commission File Number 333-171112)			3.9	Form of 1099 Service Note due 2017 filed as an Exhibit to the Company's report on Form S-8, filed on August 18, 2011, and incorporated herein by reference. (Commission File Number 333-171112)
3.10	Form of 1099 Service Note due 2018 filed as an Exhibit to the Company's report on Form S-8, filed on August 18, 2011, and incorporated herein by reference. (Commission File Number 333-171112)			3.10	Form of 1099 Service Note due 2018 filed as an Exhibit to the Company's report on Form S-8, filed on August 18, 2011, and incorporated herein by reference. (Commission File Number 333-171112)
3.11	Indefinite Asset of June 21, 2011, among the Company, the Subsidiary, Operating, and the Trustee filed as an Exhibit to the Company's report on Form S-8, filed on August 18, 2011, and incorporated herein by reference. (Commission File Number 333-171112)			3.11	Indefinite Asset of June 21, 2011, among the Company, the Subsidiary, Operating, and the Trustee filed as an Exhibit to the Company's report on Form S-8, filed on August 18, 2011, and incorporated herein by reference. (Commission File Number 333-171112)

Figure 20. An example of file from Fintabnet-test [4] dataset. Reading order of images proceeds from left to right and then to the next row. First image shows the original annotation, second image shows the masked image, third image shows the approximated table grid, and fourth image shows the output of TabGuard.

Statement of Assets		Year Ended December 31		2017		2016		2015	
Accounts receivable	1,234,567	890,123	567,890	1,234,567	890,123	567,890	1,234,567	890,123	
Inventory	345,678	234,567	123,456	345,678	234,567	123,456	345,678	234,567	
Prepaid expenses	123,456	98,765	76,543	123,456	98,765	76,543	123,456	98,765	
Other assets	567,890	456,789	345,678	567,890	456,789	345,678	567,890	456,789	
Total Assets	2,261,591	1,679,245	1,013,567	2,261,591	1,679,245	1,013,567	2,261,591	1,679,245	
Accounts payable	1,234,567	987,654	765,432	1,234,567	987,654	765,432	1,234,567	987,654	
Accrued liabilities	345,678	234,567	123,456	345,678	234,567	123,456	345,678	234,567	
Deferred revenue	123,456	98,765	76,543	123,456	98,765	76,543	123,456	98,765	
Other liabilities	567,890	456,789	345,678	567,890	456,789	345,678	567,890	456,789	
Total Liabilities	2,261,591	1,679,245	1,013,567	2,261,591	1,679,245	1,013,567	2,261,591	1,679,245	

Figure 21. An example of file from Fintabnet-test [4] dataset. Reading order of images proceeds from left to right and then to the next row. First image shows the original annotation, second image shows the masked image, third image shows the approximated table grid, and fourth image shows the output of TabGuard.

Refining & Marketing Operating Statistics By Region		2017		2016		2015	
Refined Petroleum (thousands of barrels per day)	1,234,567	1,234,567	1,234,567	1,234,567	1,234,567	1,234,567	1,234,567
Crude oil refined	456,789	456,789	456,789	456,789	456,789	456,789	456,789
Other refinery and hydrocarbon	777,778	777,778	777,778	777,778	777,778	777,778	777,778
Loss	(123,456)	(123,456)	(123,456)	(123,456)	(123,456)	(123,456)	(123,456)
Net Crude Oil Throughput (million gallons)	1,234,567	1,234,567	1,234,567	1,234,567	1,234,567	1,234,567	1,234,567
W-11 Feed Crude Oil Throughput (million gallons)	456,789	456,789	456,789	456,789	456,789	456,789	456,789
Refined Product Yields (thousands of barrels per day)	1,234,567	1,234,567	1,234,567	1,234,567	1,234,567	1,234,567	1,234,567
Gasoline	345,678	345,678	345,678	345,678	345,678	345,678	345,678
Jet fuel	234,567	234,567	234,567	234,567	234,567	234,567	234,567
Propane	123,456	123,456	123,456	123,456	123,456	123,456	123,456
Other products	510,866	510,866	510,866	510,866	510,866	510,866	510,866
Heavy fuel oil	12,345	12,345	12,345	12,345	12,345	12,345	12,345
Asphalt	12,345	12,345	12,345	12,345	12,345	12,345	12,345
Other	12,345	12,345	12,345	12,345	12,345	12,345	12,345
Refinery Direct Operating Costs (dollars per barrel)	\$ 1.47	\$ 1.47	\$ 1.47	\$ 1.47	\$ 1.47	\$ 1.47	\$ 1.47
Refined petroleum and major maintenance	1.23	1.23	1.23	1.23	1.23	1.23	1.23
Depreciation and amortization	0.15	0.15	0.15	0.15	0.15	0.15	0.15
Other manufacturing	0.09	0.09	0.09	0.09	0.09	0.09	0.09
Total	1.47	1.47	1.47	1.47	1.47	1.47	1.47

Figure 22. An example of file from Fintabnet-test [4] dataset. Reading order of images proceeds from left to right and then to the next row. First image shows the original annotation, second image shows the masked image, third image shows the approximated table grid, and fourth image shows the output of TabGuard.

	Page No.		Page No.
Management's Report on Internal Control Over Financial Reporting	39	Management's Report on Internal Control Over Financial Reporting	39
Reports of Independent Registered Public Accounting Firm	40	Reports of Independent Registered Public Accounting Firm	40
Consolidated Statements of Operations for the years ended December 31, 2016, 2015 and 2014	42	Consolidated Statements of Operations for the years ended December 31, 2016, 2015 and 2014	42
Consolidated Statements of Comprehensive Income for the years ended December 31, 2016, 2015 and 2014	43	Consolidated Statements of Comprehensive Income for the years ended December 31, 2016, 2015 and 2014	43
Consolidated Balance Sheets at December 31, 2016 and 2015	44	Consolidated Balance Sheets at December 31, 2016 and 2015	44
Consolidated Statements of Cash Flows for the years ended December 31, 2016, 2015 and 2014	45	Consolidated Statements of Cash Flows for the years ended December 31, 2016, 2015 and 2014	45
Consolidated Statements of Shareholders' Equity for the years ended December 31, 2016, 2015 and 2014	46	Consolidated Statements of Shareholders' Equity for the years ended December 31, 2016, 2015 and 2014	46
Notes to Consolidated Financial Statements	47	Notes to Consolidated Financial Statements	47
Selected Financial Data (Unaudited)	73	Selected Financial Data (Unaudited)	73
Quarterly Data and Market Price Information (Unaudited)	74	Quarterly Data and Market Price Information (Unaudited)	74

Figure 23. An example of file from Fintabnet-test [4] dataset. Reading order of images proceeds from left to right and then to the next row. First image shows the original annotation, second image shows the masked image, third image shows the approximated table grid, and fourth image shows the output of TabGuard.

Millions of dollars except per share	Year ended December 31				
	2017	2016	2015	2014	2013
Revenue	\$ 20,620	\$ 15,887	\$ 23,633	\$ 32,870	\$ 29,402
Operating income (loss)	1,362	(6,778)	(165)	5,097	3,138
Income (loss) from continuing operations	(449)	(5,767)	(662)	3,437	2,116
Basic income (loss) per share from continuing operations	(0.51)	(6.69)	(0.78)	4.05	2.35
Diluted income (loss) per share from continuing operations	(0.51)	(6.69)	(0.78)	4.03	2.33
Cash dividends per share	0.72	0.72	0.72	0.63	0.525
Net working capital	5,915	7,654	14,733	8,781	8,678
Total assets	25,085	27,000	36,942	32,165	29,223
Long-term debt	10,430	12,214	14,687	7,765	7,816
Total shareholders' equity	8,349	9,448	15,495	16,298	13,615
Capital expenditures	1,373	798	2,184	3,283	2,934

Figure 24. An example of file from Fintabnet-test [4] dataset. Reading order of images proceeds from left to right and then to the next row. First image shows the original annotation, second image shows the masked image, third image shows the approximated table grid, and fourth image shows the output of TabGuard.

Millions of dollars	Effect on	
	Pretax Pension Expense in 2009	Pension Benefit Obligation at December 31, 2009
25-basis-point decrease in discount rate	\$ 1	\$ 35
25-basis-point increase in discount rate	\$ (1)	\$ (33)
25-basis-point decrease in expected long-term rate of return	\$ 1	NA
25-basis-point increase in expected long-term rate of return	\$ (1)	NA

Figure 25. An example of file from Fintabnet-test [4] dataset, where our model failed to capture near accurate structure. Reading order of images proceeds from left to right and then to the next row. First image shows the original annotation, second image shows the masked image, third image shows the approximated table grid, and fourth image shows the output of TabGuard.

Millions of dollars	Company Shareholders' Equity					Total	
	Common Shares	Paid-in Capital in Excess of Par Value	Treasury Stock	Retained Earnings	Accumulated Other Comprehensive Income (Loss)		Noncontrolling Interest in Consolidated Subsidiaries
Balance at December 31, 2006	\$ 2,650	\$ 1,689	\$ (1,577)	\$ 5,051	\$ (437)	\$ 69	\$ 7,445
Adoption of new accounting standard	-	63	-	(43)	-	-	20
Adjusted Balance at December 31, 2006	\$ 2,650	\$ 1,752	\$ (1,577)	\$ 5,008	\$ (437)	\$ 69	\$ 7,465
Cash dividends paid	-	-	-	(14)	-	-	(14)
Stock plans	7	23	130	-	-	-	160
Common shares purchased	-	-	(1,374)	-	-	-	(1,374)
Tax benefit from exercise of options and restricted stock	-	29	-	-	-	-	29
Distributions to noncontrolling interest holders	-	-	-	-	-	(5)	(5)
Other transactions with shareholders	-	-	-	(4)	-	(21)	(25)
Total dividends and other transactions with shareholders	7	52	(1,244)	(18)	-	(26)	(1,529)
Shares exchanged in KBR, Inc. exchange offer	-	-	(2,809)	-	-	-	(2,809)
Adoption of new accounting standard	-	-	-	(30)	-	-	(30)
Comprehensive income (loss):							
Net income	-	-	-	3,486	-	50	3,536
Other comprehensive income (loss):							
Cumulative translation adjustment	-	-	-	-	1	-	1
Realization of translation gains included in net income	-	-	-	-	(24)	-	(24)
Defined benefit and other postretirement plans adjustments:							
Prior service cost:							
Plan amendment	-	-	-	(2)	-	-	(2)
Settlements/curtailments	-	-	-	5	-	-	5
Actuarial gain (loss):							
Net gain	-	-	-	105	-	-	105
Amortization of net loss	-	-	-	14	-	-	14
Settlements/curtailments	-	-	-	7	-	-	7
Tax effect on defined benefit and postretirement plans	-	-	-	(45)	-	-	(45)
KBR, Inc. separation	-	-	-	271	-	-	271
Defined benefit and other postretirement plans, net	-	-	-	355	-	-	355
Net unrealized gains on investments, net of tax provision of \$0	-	-	-	-	1	-	1
Total comprehensive income	-	-	-	3,486	333	50	3,869
Balance at December 31, 2007	\$ 2,657	\$ 1,804	\$ (5,630)	\$ 8,146	\$ (104)	\$ 93	\$ 6,966

Millions of dollars	Company Shareholders' Equity					Total	
	Common Shares	Paid-in Capital in Excess of Par Value	Treasury Stock	Retained Earnings	Accumulated Other Comprehensive Income (Loss)		Noncontrolling Interest in Consolidated Subsidiaries
Balance at December 31, 2006	\$ 2,650	\$ 1,689	\$ (1,577)	\$ 5,051	\$ (437)	\$ 69	\$ 7,445
Adoption of new accounting standard	-	63	-	(43)	-	-	20
Adjusted Balance at December 31, 2006	\$ 2,650	\$ 1,752	\$ (1,577)	\$ 5,008	\$ (437)	\$ 69	\$ 7,465
Cash dividends paid	-	-	-	(14)	-	-	(14)
Stock plans	7	23	130	-	-	-	160
Common shares purchased	-	-	(1,374)	-	-	-	(1,374)
Tax benefit from exercise of options and restricted stock	-	29	-	-	-	-	29
Distributions to noncontrolling interest holders	-	-	-	-	-	(5)	(5)
Other transactions with shareholders	-	-	-	(4)	-	(21)	(25)
Total dividends and other transactions with shareholders	7	52	(1,244)	(18)	-	(26)	(1,529)
Shares exchanged in KBR, Inc. exchange offer	-	-	(2,809)	-	-	-	(2,809)
Adoption of new accounting standard	-	-	-	(30)	-	-	(30)
Comprehensive income (loss):							
Net income	-	-	-	3,486	-	50	3,536
Other comprehensive income (loss):							
Cumulative translation adjustment	-	-	-	-	1	-	1
Realization of translation gains included in net income	-	-	-	-	(24)	-	(24)
Defined benefit and other postretirement plans adjustments:							
Prior service cost:							
Plan amendment	-	-	-	(2)	-	-	(2)
Settlements/curtailments	-	-	-	5	-	-	5
Actuarial gain (loss):							
Net gain	-	-	-	105	-	-	105
Amortization of net loss	-	-	-	14	-	-	14
Settlements/curtailments	-	-	-	7	-	-	7
Tax effect on defined benefit and postretirement plans	-	-	-	(45)	-	-	(45)
KBR, Inc. separation	-	-	-	271	-	-	271
Defined benefit and other postretirement plans, net	-	-	-	355	-	-	355
Net unrealized gains on investments, net of tax provision of \$0	-	-	-	-	1	-	1
Total comprehensive income	-	-	-	3,486	333	50	3,869
Balance at December 31, 2007	\$ 2,657	\$ 1,804	\$ (5,630)	\$ 8,146	\$ (104)	\$ 93	\$ 6,966

Figure 26. An example of file from Fintabnet-test [4] dataset, where our model failed to capture near accurate structure. Reading order of images proceeds from left to right and then to the next row. First image shows the original annotation, second image shows the masked image, third image shows the approximated table grid, and fourth image shows the output of TabGuard.

Millions of dollars	Years ended December 31			Total
	2004	2003	2002	
Depreciation, depletion, and amortization:				
Production Optimization	\$ 115	\$ 104	\$ 99	
Fluid Systems	60	50	48	
Drilling and Formation Evaluation	115	144	137	
Digital and Consulting Solutions	75	77	112	
Shared Energy Services	91	92	79	
Total Energy Services Group	456	467	475	
Government and Infrastructure	27	22	11	
Energy and Chemicals	11	16	17	
Shared KBR	15	12	1	
Total KBR	53	50	29	
General corporate	-	1	1	
Total	\$ 509	\$ 518	\$ 505	
Bad assets:				
Production Optimization	\$ 1,754	\$ 1,659	\$ 1,444	
Fluid Systems	1,045	1,030	890	
Drilling and Formation Evaluation	960	1,074	1,163	
Digital and Consulting Solutions	768	794	1,320	
Shared Energy Services	1,021	1,240	1,187	
Total Energy Services Group	5,548	5,797	5,944	
Government and Infrastructure	3,309	2,758	784	
Energy and Chemicals	1,656	2,078	2,055	
Shared KBR	198	246	265	
Total KBR	5,163	5,082	3,104	
General corporate	5,085	4,620	3,796	
Total	\$15,796	\$15,499	\$12,844	

Figure 27. An example of file from Fintabnet-test [4] dataset, where our model failed to capture near accurate structure. Reading order of images proceeds from left to right and then to the next row. First image shows the original annotation, second image shows the masked image, third image shows the approximated table grid, and fourth image shows the output of TabGuard.

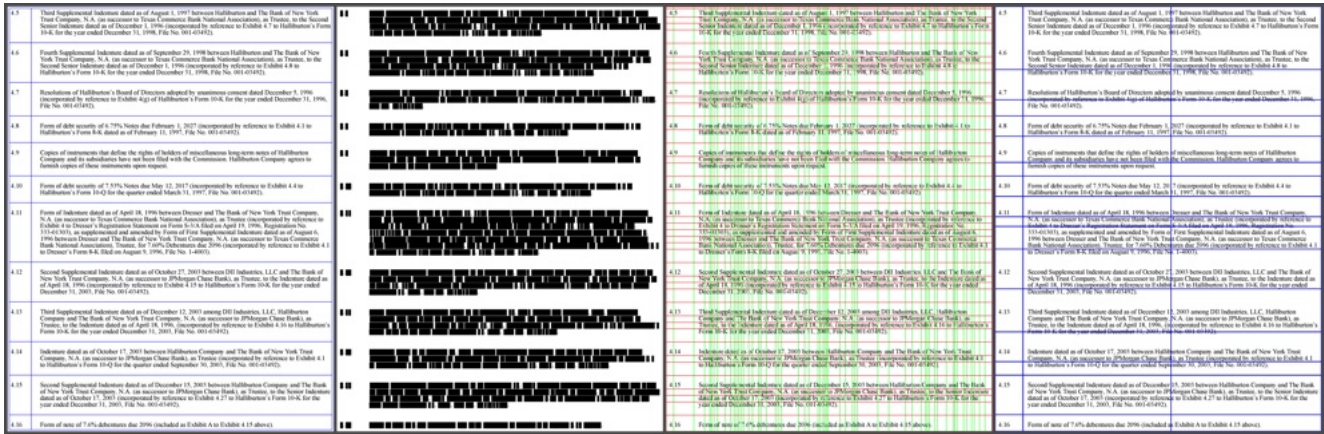


Figure 28. An example of file from Fintabnet-test [4] dataset, where our model failed to capture near accurate structure. Reading order of images proceeds from left to right and then to the next row. First image shows the original annotation, second image shows the masked image, third image shows the approximated table grid, and fourth image shows the output of TabGuard.

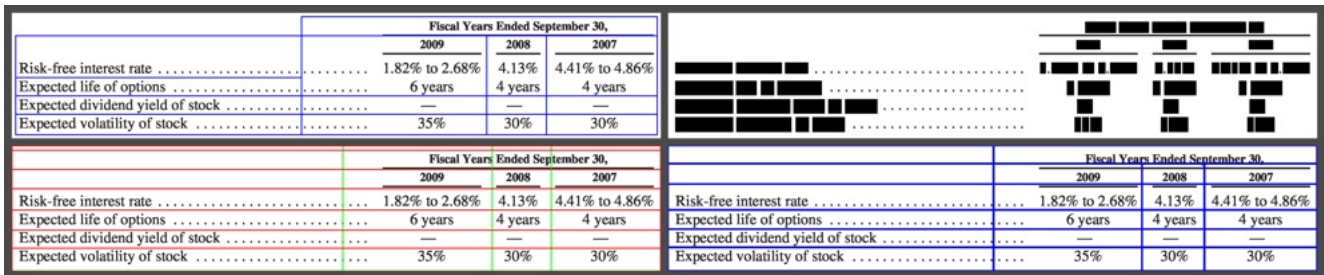


Figure 29. An example of file from Fintabnet-test [4] dataset which is incorrectly annotated. Reading order of images proceeds from left to right and then to the next row. First image shows the original annotation, second image shows the masked image, third image shows the approximated table grid, and fourth image shows the output of TabGuard.

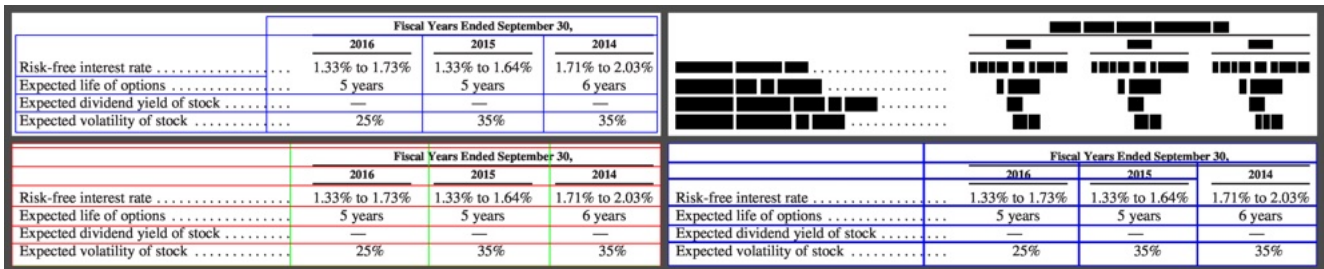


Figure 30. An example of file from Fintabnet-test [4] dataset which is incorrectly annotated. Reading order of images proceeds from left to right and then to the next row. First image shows the original annotation, second image shows the masked image, third image shows the approximated table grid, and fourth image shows the output of TabGuard.

Location	Products Manufactured	Segment	Approximate Area (Sq. Feet)
Mississauga, Canada ⁽⁶⁾	---	Consumer International	123,000
Folkestone, England ⁽⁷⁾	---	Consumer International	78,500
Revel, France	---	Consumer International	35,500
Mexico City, Mexico	---	Consumer International	27,500
Sydney, Australia	---	Consumer International	24,900
Feira de Santana, Bahia, Brazil	---	SPD	21,700
Langley, British Columbia ⁽⁸⁾	---	Consumer International	6,300
Offices			
Levallois, France	---	Consumer International	21,600
Mississauga, Canada	---	Consumer International	17,000
Folkestone, England ⁽⁹⁾	---	Consumer International	11,000
Dover, England	---	Consumer International	9,400

Location	Products Manufactured	Segment	Approximate Area (Sq. Feet)
Mississauga, Canada ⁽⁶⁾	---	Consumer International	123,000
Folkestone, England ⁽⁷⁾	---	Consumer International	78,500
Revel, France	---	Consumer International	35,500
Mexico City, Mexico	---	Consumer International	27,500
Sydney, Australia	---	Consumer International	24,900
Feira de Santana, Bahia, Brazil	---	SPD	21,700
Langley, British Columbia ⁽⁸⁾	---	Consumer International	6,300
Offices			
Levallois, France	---	Consumer International	21,600
Mississauga, Canada	---	Consumer International	17,000
Folkestone, England ⁽⁹⁾	---	Consumer International	11,000
Dover, England	---	Consumer International	9,400

Location	Products Manufactured	Segment	Approximate Area (Sq. Feet)
Mississauga, Canada ⁽⁶⁾	---	Consumer International	123,000
Folkestone, England ⁽⁷⁾	---	Consumer International	78,500
Revel, France	---	Consumer International	35,500
Mexico City, Mexico	---	Consumer International	27,500
Sydney, Australia	---	Consumer International	24,900
Feira de Santana, Bahia, Brazil	---	SPD	21,700
Langley, British Columbia ⁽⁸⁾	---	Consumer International	6,300
Offices			
Levallois, France	---	Consumer International	21,600
Mississauga, Canada	---	Consumer International	17,000
Folkestone, England ⁽⁹⁾	---	Consumer International	11,000
Dover, England	---	Consumer International	9,400

Figure 31. An example of file from Fintabnet-test [4] dataset which is incorrectly annotated. Reading order of images proceeds from left to right and then to the next row. First image shows the original annotation, second image shows the masked image, third image shows the approximated table grid, and fourth image shows the output of TabGuard.