Models	Word			Latex			Word+Latex		
	Precision	**Recall**	**F1**	**Precision**	**Recall**	**F1**	**Precision**	**Recall**	**F1**
ResNeXt-101 (Word)	0.9496	0.8388	0.8908	0.9902	0.5948	0.7432	0.9594	0.7607	0.8486
ResNeXt-152 (Word)	0.9530	0.8829	**0.9166**	0.9808	0.6890	0.8094	0.9603	0.8209	0.8851
ResNeXt-101 (Latex)	0.8288	0.9395	0.8807	0.9854	0.9760	0.9807	0.8744	0.9512	0.9112
ResNeXt-152 (Latex)	0.8259	0.9562	0.8863	0.9867	0.9754	**0.9810**	0.8720	0.9624	0.9149
ResNeXt-101 (Word+Latex)	0.9557	0.8403	0.8943	0.9886	0.9694	0.9789	0.9670	0.8817	0.9224
ResNeXt-152 (Word+Latex)	0.9540	0.8639	0.9067	0.9885	0.9732	0.9808	0.9657	0.8989	**0.9311**

Table 2: Evaluation results on Word and Latex datasets with ResNeXt-$\{101,152\}$ as the backbone networks

Models	Word	Latex	Word+Latex
Image-to-Text (Word)	**0.7507**	0.6733	0.7138
Image-to-Text (Latex)	0.4048	**0.7653**	0.5818
Image-to-Text (Word+Latex)	0.7121	0.7647	**0.7382**

Table 3: Evaluation results (BLEU) for image-to-text models on Word and Latex datasets

The evaluation results of table structure recognition are shown in Table 3. We observe that the image-to-text models also perform better on the same domain. The model trained on Word documents performs much better on the Word test set than the Latex test set and vice versa. Similarly, the model accuracy of the Word+Latex model is comparable to other models on Word and Latex domains and better on the mixed-domain dataset. This demonstrates that the mixed-domain model might generalize better in real world applications.

5.4 Analysis

For table detection, we sample some incorrect examples from evaluation data for the case study. Figure 7 gives three typical errors of detection results. The first error type is **partial-detection**, where only part of the tables can be identified and some information is missing. The second error type is **un-detection**, where some tables in the documents cannot be identified. The third error type is **mis-detection**, where figures and text blocks in the documents are sometimes identified as tables. Taking the ResNeXt-152 model for Word+Latex as an example, the number of un-detected tables is 164. Compared with ground truth tables (2,525), the un-detection rate is 6.5%. Meanwhile, the number of mis-detected tables is 86 compared with the total predicted tables being 2,450. Therefore, the mis-detection rate is 3.5%. Finally, the number of partial-detected tables is 57, leading to a partial-detection rate of 2.3%. This illustrates that there is plenty of room to improve the accuracy of the detection models, especially for un-detection and mis-detection cases.

For table structure recognition, we observe that the model accuracy reduces as the length of output becomes larger. Taking the image-to-text model for Word+Latex as an example, the number of exact match between the output and ground truth is shown in Table 4. We can see that the ratio of exact match is around 50% for the HTML sequences that are less than 40 tokens. As the number of tokens becomes larger, the ratio reduces dramatically to 8.6%, indicating that it is more difficult to recognize big and complex tables. In general, the model totally generates the correct output for 338 tables. We believe enlarging the training data will further improve the current model especially for tables with complex row and column layouts, which will be our next-step effort.

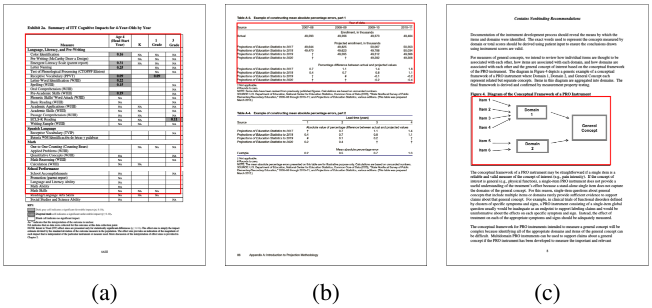
(a) (b) (c)

Figure 7: Table detection examples with (a) partial-detection, (b) un-detection and (c) mis-detection

Length	0-20	21-40	41-60	61-80	>80	All
#Total	32	293	252	145	278	1,000
#Exact match	15	169	102	28	24	338
Ratio	0.469	0.577	0.405	0.193	0.086	0.338

Table 4: Number of exact match between the generated HTML tag sequence and ground truth sequence

6 Conclusion

To empower the research of table detection and structure recognition for document analysis, we introduce the TableBank dataset, a new image-based table analysis dataset built with online Word and Latex documents. We use the Faster R-CNN model and image-to-text model as the baseline to evaluate the performance of TableBank. In addition, we have also created testing data from Word and Latex documents respectively, where the model accuracy in different domains is evaluated. Experiments show that image-based table detection and recognition with deep learning is a promising research direction. We expect the TableBank dataset will release the power of deep learning in the table analysis task, meanwhile fosters more customized network structures to make substantial advances in this task.

For future research, we will further enlarge the TableBank from more domains with high quality. Moreover, we plan to build a dataset with multiple labels such as tables, figures, headings, subheadings, text blocks and more. In this way, we