Table Structure Recognition – Work In Progress

Anonymous ICCV submission

Paper ID ****

Abstract

Add Pipeline, Network Diagram

*Information pertaining to a specific topic is often presented in a compact and structured tabular format in documents. However, extracting information from such tabular data from images of unstructured documents is a non-trivial yet vital component of any document information extraction pipeline. To aid the aforementioned task, we present a framework that facilitates table structure recognition from images of table @Andras: With additional inputs, i.e., cell-texts, their positions, contents?. @Rudra: Both the experimental regimes i.e. with and with-out additional inputs. Pipeline similar to TabStructNet. We formulate the task as predicting adjacency relationship between non-empty cells of the table and classifying the same as row or column specific relationship. To this end, we utilize rules dependent upon bounding boxes to generate graphs @Rudra: **WIP**. Our proposed framework integrates cues from image features of the table as well as the corresponding position features of non-empty cells. We validate the efficacy of the framework on multiple publicly available datasets and demonstrate how it fares compared to contemporary works.*

1. Introduction

- Overall Motivation: Ever-growing need for extracting information from images of structured/unstructured documents with heterogeneous elements such as tables, images etc.
  - Extracting information from each of these different elements presents a unique set of challenges.
  - This work presents a framework that helps facilitate information extraction from images of tables.
  - Reliance on cell bounding boxes.
  - Heterogeneous table formats
  - Failing to detect errors caused by empty cells and misalignment of cells beyond immediate neighbors
  - Partial border/lines
  - Tilt, shear, rotation in table images (scanned)
  - Text justification
  - We view this as a problem of reasoning over a graph.
- Limitations of Current Works:
  - Current Graph based works [2, 3] only consider immediate adjacency relationship between non-empty cells created using methods like kNN.
    - They fail to detect errors caused by empty cells and misalignment of cells beyond immediate neighbors [6].
    - Multiple layers of graph processing is required in them to pool global information together for classifying a given edge.
    - Such a network will not generalize well as just stacking multiple layers of graph processing doesn't work out of the box [1] and if say the architecture is carefully constructed depending upon the number of rows/cols seen during training, it might not generalize well for larger tables.
  - Since tables are heterogeneous structures, it will be useful to employ attention mechanism to focus on specific set of cells while making decision about an edge.
    - Current works do not have any such mechanism.
  - Motivation for proposed framework:
    - Applying Graph Attention (GAT-GCNs) based architecture on graphs generated using rules to learn node representation as part of a pair-wise node classification problem.

- · Classifying node pairs as being part of same-row, same-col etc.
- ∗ Rules help us connect cells beyond immediate neighbors for graph processing:
  - · Row and Col specific rules generating row and col adjacency matrix.
  - · **Naive Gaussian Graph Generation Process:** Consider the cell-text (nodes of our graphs) under consideration. We apply a 1D Gaussian function along the y-axis and x-axis of this cell text to obtain the set of cells that are potentially part of the same row and col generating row and col specific adjacency matrices respectively. These adjacency matrices are combined for multi-class and multi-task based training.
- ∗ Few layers of graph processing required as relevant information not multiple hops away even for larger tables.
- ∗ Rules are independent of the size of the table.

2. Related Work

3. Methodology

4. Experiments and Results

4.1. Datasets

- SciTSR, Synthetic
- ICDAR 2013, 2019, 2021 *@Andras: If 2021 refers to the ICDAR 2021 IBM Challenge then it is not possible; inputs are table images only, also, test dataset might not be available*

Baselines for comparison:

- GFTE *@Andras: I believe that it is only an arXiv paper..*
- Rethinking *@Andras: TF-based, results only on their synthetic dataset*
- Table Structure Recognition using Top-Down and Bottom-Up Cues
- Adobe paper of ICDAR 2019 (Chris Tenyonson's work)
- Work reporting only on ICDAR 2013 (DeepDeSRT, etc.) *@Andras: DeepDeSRT results are reported on some random subset of ICDAR 2013, model was not released*

4.2. Metrics for Evaluation

- F1, Precision, Recall of pairwise adjacency relations between non-empty cells
- Tree-Edit-Distance base Similarity (TEDS): @Gautam, @CK This metric is only applicable when we present the final output in the form of a tree which our current setup doesn't. We can think about how to adapt this, which I think will be a good idea. *@Andras: TEDS includes structure AND content!!*

4.3. Ablation Study

- Ablation of features
- Since we are proposing to train CNN features using self-supervised learning as a pre-training step. That is one of our ablation study.

4.4. Miscellaneous

Improving results of Graph-Rules work on ICDAR 2013 dataset:

- Train on larger or complex dataset if needed
  - Rethinking work doesn't report Precision, Recall and F1. Result of TabStructNet and Rethinking work mentioned in TabStructNet are suspicious.
- Find variance values that fit ICDAR 2013 dataset better

4.5. Results

5. Conclusion

6. Final Copy

References

[1] Guohao Li, Matthias Müller, Ali Thabet, and Bernard Ghanem. Deepgcns: Can gcns go as deep as cnns? In *The IEEE International Conference on Computer Vision (ICCV)*, 2019.

[2] Yiren Li, Zheng Huang, Junchi Yan, Yi Zhou, Fan Ye, and Xianhui Liu. Gfte: Graph-based financial table extraction, 2020.

[3] Shah Rukh Qasim, Hassan Mahmood, and Faisal Shafait. Rethinking table recognition using graph neural networks, 2019.

[4] Sachin Raja, Ajoy Mondal, and C. V. Jawahar. Table structure recognition using top-down and bottom-up cues, 2020.

[5] Chris Tensmeyer, Vlad I. Morariu, Brian Price, Scott Cohen, and Tony Martinez. Deep splitting and merging for table structure decomposition. In *2019 International Conference on Document Analysis and Recognition (ICDAR)*, pages 114–121, 2019.

[6] Xu Zhong, Elaheh ShafieiBavani, and Antonio Jimeno-Yepes. Image-based table recognition: data, model, and evaluation. *CoRR*, abs/1911.10683, 2019.

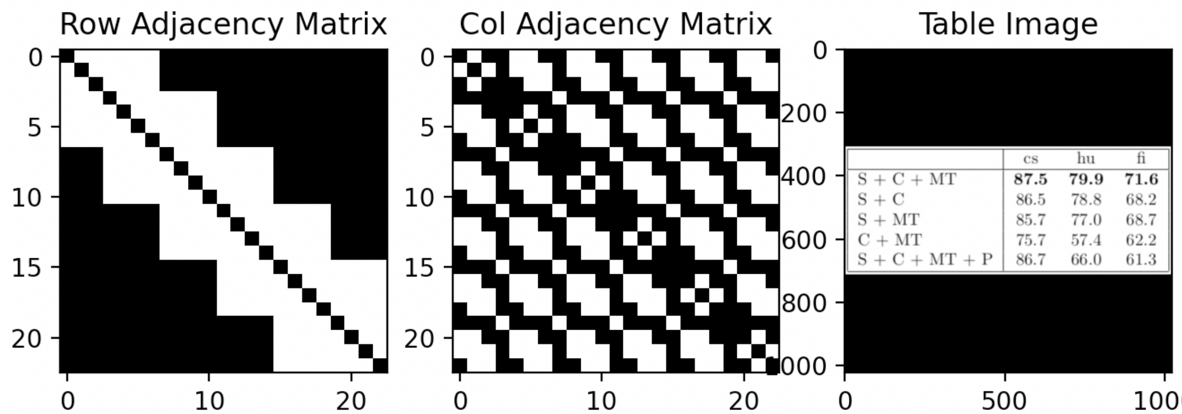

Figure 1. Row and Column specific adjacency matrices of the cell-text bounding boxes generated for the given table image by applying the naive Gaussian graph generation rule upon the centroid of the cell-text bounding boxes.

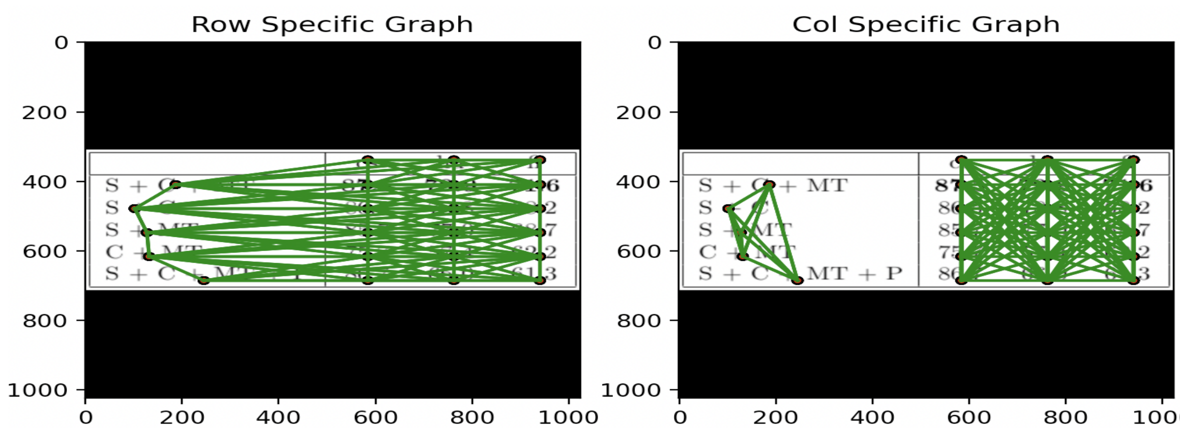

Figure 2. Row and Column specific graphs for the generated adjacency matrices shown in Fig 1

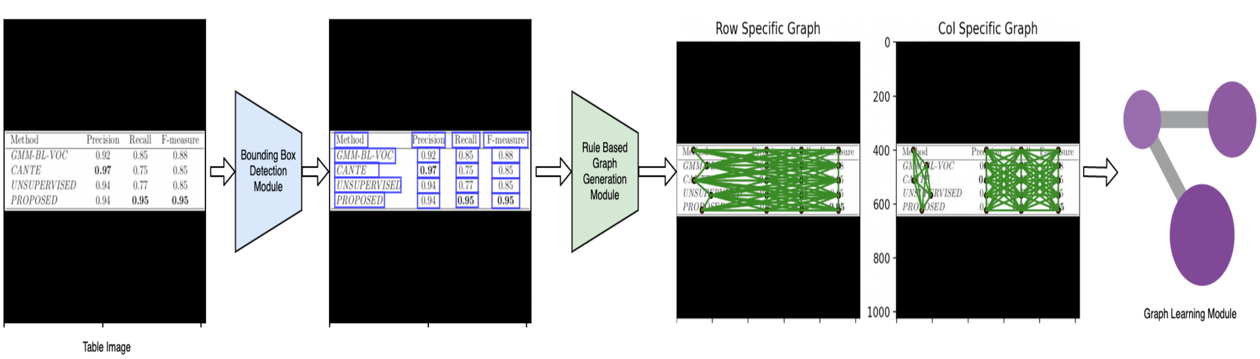

Figure 3. Pipeline depicting processing modules applied on the input image of the table to deduce its structure information.

Rule-Based Approach (Only Position Features)							
Rule:	Naive Gaussian on Centroid						
Datasets:	SciTSR				ICDAR 2013		
Metrics:	Precision TSR	Recall TSR	F1 TSR	Accuracy (row, col)	Precision	Recall	F1
GR-GCN Single-rel	0.920	0.915	0.917	0.970 (0.985, 0.955)	0.912	0.908	0.91
GR-GCN Multi-class							
GR-GCN Multi-task							
GR-GAT(variant) Single-rel							
GR-GAT Multi-rel							
GR-GAT Multi-task							
Related Works							
GFTE	-	-	-	0.955 (0.964, 0.946)	-	-	-
TabStructNet				-	0.976	0.985	0.981
Split-PDF				-	0.920	0.913	0.916
Split-PDF + Heuristic				-	0.959	0.946	0.953

Table 1. Results of different models on test sets. GFTE results are for re-trained models after fixing issues of original implementation. TabStructNet results are from [4]. Split-PDF results are from [5]. GFTE results take into account image and text features along-with position features. TabStructNet and Split-PDf ingests image features for its reasoning as well. **Add Precision, Recall and F1 for Row and Col adjacency, Qualitative Results (Multi-row/col examples, lop-sided cell-text boxes)**