MediTables: A New Dataset and Deep Network for Multi-Category Table Localization in Medical Documents

by

Akshay Praveen Deshpande, Vaishnav Rao Potlapalli, Ravi Kiran Sarvadevabhatla

in

IAPR International Workshop on Graphics Recognition (GREC 2021)
: 1
-14

Report No: IIIT/TR/2021/-1

Centre for Visual Information Technology
International Institute of Information Technology
Hyderabad - 500 032, INDIA
September 2021

MediTables: A New Dataset and Deep Network for Multi-Category Table Localization in Medical Documents

Akshay Praveen Deshpande[0000−0002−4632−5535], Vaishnav Rao Potlapalli[0000−0003−0923−6645], and Ravi Kiran Sarvadevabhatla (✉)[0000−0003−4134−1154]

Centre for Visual Information Technology
International Institute of Information Technology, Hyderabad (IIIT-H)
Hyderabad, INDIA 500032
akshaydp1995@gmail.com,pvaishnav2718@gmail.com,ravi.kiran@iiit.ac.in
https://github.com/atmacvit/meditables

Abstract. Localizing structured layout components such as tables is an important task in document image analysis. Numerous layout datasets with document images from various domains exist. However, healthcare and medical documents represent a crucial domain that has not been included so far. To address this gap, we contribute MediTables, a new dataset of 200 diverse medical document images with multi-category table annotations. Meditables contains a wide range of medical document images with variety in capture quality, layouts, skew, occlusion and illumination. The dataset images include pathology, diagnostic and hospital-related reports. In addition to document diversity, the dataset includes implicitly structured tables that are typically not present in other datasets. We benchmark state of the art table localization approaches on the MediTables dataset and introduce a custom-designed U-Net which exhibits robust performance while being drastically smaller in size compared to strong baselines. Our annotated dataset and models represent a useful first step towards the development of focused systems for medical document image analytics, a domain that mandates robust systems for reliable information retrieval. The dataset and models can be accessed at https://github.com/atmacvit/meditables .

Keywords: Document Analysis · Table Localization · Healthcare · Medical · Semantic Segmentation · Instance Segmentation

1 Introduction

Document tables have been an efficient and effective technique for communicating structured information. With the advent of digital media, most document tables exist in PDF documents. Consequently, there have been efforts to develop algorithms for machine-based detection and understanding of tabular content from such media.

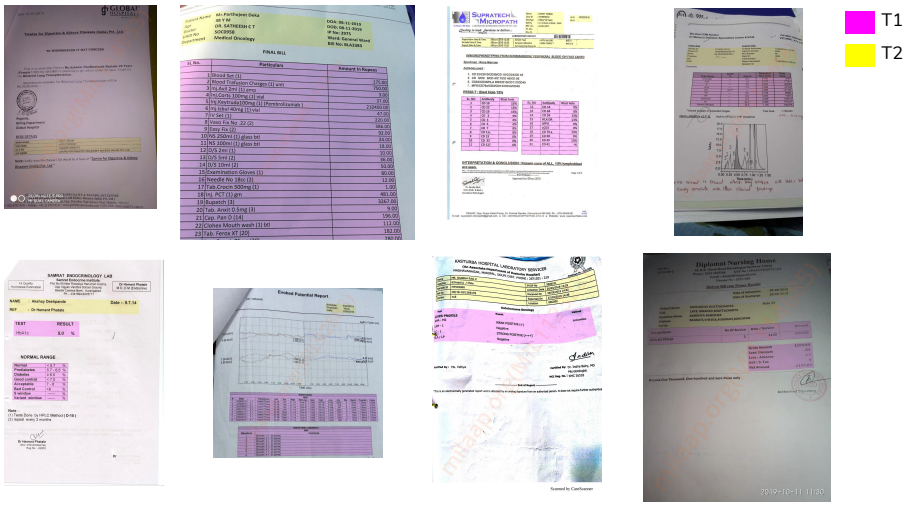

Fig. 1: Samples from our MediTables dataset. T1 and T2 represent two different kinds of table annotations. The diversity of the medical documents and table configurations is visible in the figure.

To understand and develop efficient, accurate algorithms for table localization and understanding, diverse datasets of document images were created and have been made available to the research community. In general, tables appear in varying formats and layouts which thwart heuristic approaches. Of late, deep learning [12] has proven to be a powerful mechanism to obtain state of the art results on many computer vision tasks, including table understanding.

While the existing datasets cover a number of domains, documents related to healthcare and medical domain are conspicuously absent. We seek to address this gap by contributing a new, annotated dataset. The documents in this dataset pose challenges not encountered in other datasets. Thus, this dataset adds to the diversity of the dataset pool.

Most of the available document datasets tend to contain similar levels of illumination and capture quality. Due to qualitative differences between existing datasets and medical documents, we also discover that pre-trained, deep-learning based table localization models trained on existing datasets do not generalize sufficiently on documents from the medical domain. Therefore, we also introduce a customized deep network for table localization in medical domain documents.

Specifically, we make the following contributions:

- A medical document image dataset called MediTables with annotations for two different types of tables.
- A deep learning model for table localization in medical document images.

The dataset and models can be accessed at https://github.com/atmacvit/meditables .

2 Related Work

Approaches for table localization characterize the problem either as a detection problem (i.e., identify axis-aligned bounding boxes for tables) or as a segmentation problem (i.e., obtain pixel-level labeling for tables). We review the literature related to these two approaches below.

Table Detection: There has been substantial work done in the wider context of table detection in document images and PDFs. Many of the earlier approaches were heuristic-based approaches. Ha et al. [7] proposed Recursive X-Y cut algorithm which recursively decomposes the document in blocks which are then used to build a X-Y tree which is later used for segmentation. Kieninger et al. [11] proposed a system called T-Recs which uses bottom up clustering of word entities and word geometries into blocks for segmentation of tables. Yildiz et al. [25] introduced a system known as pdf2table which uses a tool called pdf2html to convert the PDF file to its XML counterpart containing information regarding absolute location of text chunks. This XML along with certain heuristic rules is then used to perform table localization. Fang et al. [1] also showed a method which works on PDF documents. They use a four stepped approach for performing table detection. First, they parse PDF files and perform layout analysis. Separators and delimiting information is then mined to localize tables.

Hao et al. [8] propose a method where table regions are initially selected based on some heuristic rules and later, a convolutional neural network is trained. This is used to classify the selected regions into tables and non-tables. Huang et al. [9] use a modified YOLO framework for table localization with post processing step for additional performance improvement. Schreiber et al. [20] use a fine-tuned version of popular object detection framework, Faster RCNN [4], to detect tables in document images. Siddiqui et al. [21] propose a framework called DeCNT. In this approach, they combine a deformable CNN model with RCNN or an FCN instead of a conventional CNN. Gilani et al. [3] perform various distance transforms on the original image to create an alternative representation which is subsequently fed into an Faster R-CNN model for detecting tables.

Table Segmentation: Yang et al. [24] propose a multi-modal Fully Convolutional Network which segments documents images into various page elements (text, charts, tables) using both image information as well as underlying text in the image. Kavasidis et al.[10] model table localization as a semantic segmentation problem and use a Fully Convolutional Network pretrained on saliency detection datasets to develop visual cues similar to those of tables and graphs. The predictions of the network are further refined using Conditional Random Fields (CRFs). CascadeTabNet [14] is an Mask-RCNN based network trained to localize table regions in reasonably structured documents.

3 MediTables dataset

We introduce MediTables, a dataset of 200 multi-class annotated medical document images. The document images were scraped from various sources on the

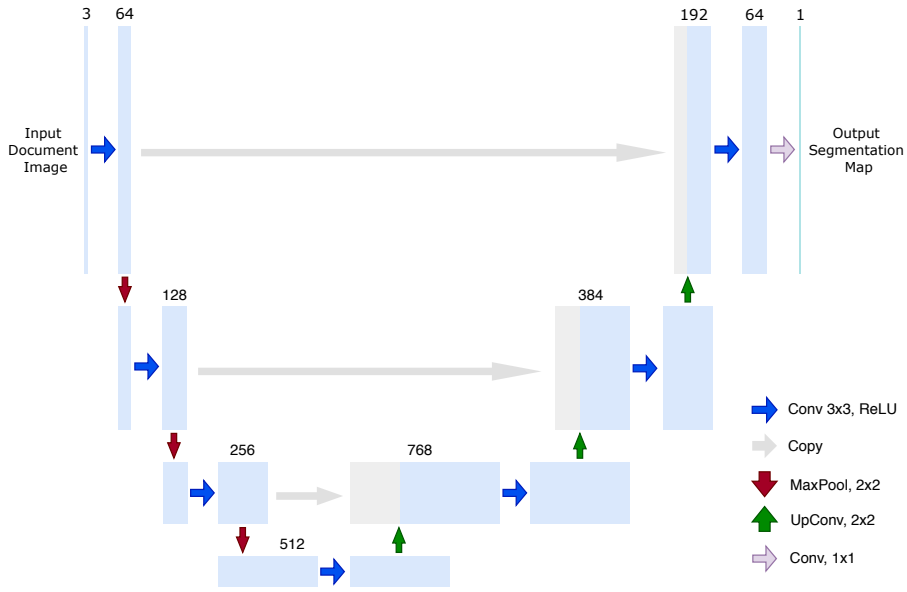

Fig. 2: The architecture for our modified U-Net.

Table 1: Table coverage by type in MediTables.

	# of documents				# of tables			
Type	Total	Train	Validation	Test	Total	Train	Validation	Test
T1 and T2	79	60	7	12	–	–	–	–
T1 only	73	43	15	15	190	126	29	35
T2 only	48	27	8	13	140	100	15	25
Total	200	130	30	40	330	226	44	60

internet. Our dataset contains a wide range of medical document images with variety in capture quality, layouts, skew, occlusion and illumination. They are a good representation of prevalent healthcare, medical images such as pathology, diagnostic and hospital-related reports. There are two kinds of annotated tables in our dataset (see Figure 1):

- The first kind of table, hereafter referred to as T1, are tables which follow a conventional layout and tend to have some sort of demarcation between rows or between columns.
- The other kind of table, hereafter referred to as T2, consists of formatted data in key-value format which are usually found for fields such as names, identification numbers, addresses, age, etc.

The inclusion of such annotations can facilitate efficient retrieval of crucial meta-information from medical document images. Additional statistics related to the dataset can be seen in Table 1.

4 The modified U-Net Deep Network

Due to the informal nature of document capture and consequent distortions induced, table layouts in our dataset are non-rectangular. Therefore, we model the task as a semantic segmentation problem and use a custom-designed U-Net [18] to localize tables. We chose a U-Net based model due to its success for segmentation tasks trained on small datasets. U-Net uses skip connections, which improves gradient flow and allows stable weight updates.

The original U-Net [18] has four skip connections across the network and requires cropping to meet the size demands for establishing skip connections. We modify the original U-Net to exclude the cropping in the cropping and copying step. This is because the skip connections are performed across layers that have the same spatial size. Thus, our modified U-Net has only three skip connections. Another difference is that we use a single convolution layer per down-sampling layer compared to double convolutions used in the original U-Net since we found performance to be empirically better with this design choice.

The model consists of a contraction and an expansion section. The contraction section of the model consists of 4 convolutional layers with 3×3 filters of $64, 128, 256$ and 512 channels successively. Each layer is followed by a Rectified Linear Unit (ReLU) activation layer. The convolutions are all single stride with a padding size of 1. The outputs of each convolution layer are max-pooled using a 2×2 kernel. The expansion section consists of 3 convolutional layers which perform upsampling using a 2×2 kernel with a padding of 1. After each upsampling step, features from the contraction section that have the same spatial dimensions as the current set of features are concatenated. The output, which has the same spatial dimensions as the input image, is obtained by applying 1×1 convolution filters on the feature output of the expansion section.

5 Experiment Setup

For our experiments, we consider multiple popular table localization datasets as described below.

5.1 Datasets

Marmot: The Marmot table recognition dataset consists of 2000 PDF document images with diverse page layouts and tabular formats. The Marmot layout analysis of fixed layout documents dataset consists of 244 clean document images and comprises of 17 labels for fragments in a document. From this, we selected 400 relevant images with tables.

Table 2: Performance comparison of original U-Net and modified U-Net evaluated on validation set of MediTables.

Model	Loss	IOU (%)	PPA (%)	F1 (%)
Original U-Net	$\mathcal{L}_{BCE}$	71.57	86.31	81.26
	$\mathcal{L}_{IoU}$	74.78	88.74	85.10
	$\mathcal{L}_{BCE} + \mathcal{L}_{IoU}$	**76.21**	89.39	86.20
Modified U-Net	$\mathcal{L}_{BCE}$	73.20	88.20	83.90
	$\mathcal{L}_{IoU}$	75.72	89.60	85.40
	$\mathcal{L}_{BCE} + \mathcal{L}_{IoU}$	75.81	**90.06**	**87.08**

UNLV: The UNLV dataset [22] contains 2889 scanned document images from sources such as newspapers and business letters. The resolution for such images are 200 to 300 DPI. We use the subset of 427 images containing tables for experiments.

UW3: The dataset consists of 1600 scanned skew-corrected document images. We have selected 120 document images, which contain atleast one table in them.

ICDAR Datasets: We use 124 documents from the ICDAR 2013 table detection competition [5]. Additionally, we used 549 document images with tables from ICDAR 2017 Page Object Detection Dataset [2].

TableBank: Recently, researchers from Beihang University and Microsoft Research Asia collected the largest document image based dataset with tables, TableBank [13]. It consists of 417,000 labeled tables and clean source documents.

To augment the datasets, we performed various standard augmentations on images such as Gaussian blurring, rotation, salt and pepper noise, Poisson noise and affine transformation. The augmentations were randomly applied to obtain a dataset consisting of 52,482 images.

5.2 Training and Implementation Details

To begin with, we trained our modified U-Net model using combined data from four existing datasets (Section 5.1). The resulting model was fine-tuned on the training set of 130 images sourced from our MediTables dataset.

All images and corresponding label map targets were resized to 512×512. For the optimization of our network, we used the popular Adam optimizer with a learning rate of 5×10^{-4}, with a corresponding mini-batch size of 16. The training was conducted in two phases. In the first phase of training, we used per-pixel binary cross-entropy loss ($\mathcal{L}_{BCE} = -y * log(\bar{y}) - (1-y) * log(1-\bar{y})$) where y is the ground-truth label, $\bar{y}$ is the prediction), for 15 epochs. In the second phase (i.e., 16th epoch onwards), we included the logarithmic version of IoU loss [16]

Table 3: Performance comparison between pre-trained and fine-tuned models.

(a) Performance of models trained only on existing document datasets and evaluated on test set of MediTables (PPA is not defined for detection models).

Model	IOU (%)	PPA (%)	F1 (%)
TableBank [13]	89.27 ± 18.21	NA	90.26 ± 17.24
YOLO-v3 [17]	19.85 ± 08.10	NA	17.44 ± 07.09
pix2pixHD [23]	21.06 ± 02.44	90.24 ± 06.13	42.37 ± 06.32
CascadeTabNet [14]	83.08 ± 19.00	95.70 ± 07.00	93.05 ± 01.00
Modified U-Net (Ours)	21.32 ± 16.16	71.14 ± 22.03	32.44 ± 21.10

(b) Performance of models pre-trained on existing document datasets and fine-tuned on MediTables (PPA is not defined for detection models).

Model	IOU (%)	PPA (%)	F1 (%)
TableBank [13]	95.15 ± 04.84	NA	97.10 ± 01.36
YOLO-v3 [17]	45.48 ± 25.47	NA	45.34 ± 11.41
pix2pixHD [23]	88.16 ± 06.91	97.61 ± 01.51	97.63 ± 00.74
CascadeTabNet [14]	89.61 ± 17.82	97.16 ± 05.26	95.07 ± 08.00
Modified U-Net (Ours)	$\mathbf{96.77 \pm 02.03}$	$\mathbf{99.51 \pm 00.21}$	$\mathbf{99.48 \pm 00.12}$

$$\mathcal{L}_{IoU} = -ln\left(\frac{X \cap \hat{X}}{X \cup \hat{X}}\right) \tag{1}$$

where $\hat{X}$ is the predicted image mask and X is the corresponding ground-truth label mask. In contrast with per-pixel cross-entropy loss, IoU loss optimizes for the table regions in a more direct manner and turns out to be crucial for overall performance, as we shall see shortly. The final loss $\mathcal{L}$, used during the second phase of training, is a weighted combination of the aforementioned losses, i.e. $\mathcal{L} = \lambda_1 \mathcal{L}_{BCE} + \lambda_2 \mathcal{L}_{IoU}$ where $\lambda_1 = 1, \lambda_2 = 20$. The network is trained for a total of 58 epochs.

Finally, we combined training, validation sets and re-trained the model with the hyper-parameters and stopping criteria determined from the validation set experiments. The final model is evaluated on a disjoint test set (40 images). The entire training was performed using four Nvidia GeForce GTX 1080 Ti GPUs.

6 Experiments and Analysis

For all experiments, we compare performance using the standard measures – Intersection over Union (IoU) [19], Per pixel Average (PPA) [15], F-1 (Dice) score [6] – averaged over the evaluation set. The IoU is calculated corresponding to the tabular part of the document image, and the resulting nearest tabular mask from the model's output. The F-1 score is calculated using the conventional Dice coefficient formula for tabular regions of the document and the corresponding model outputs.

Modified v/s Original U-Net: To begin with, we compared the performance of original and modified U-Net. As Table 2 shows, our modified U-Net has a small but significant performance advantage over the original version, justifying its choice. The relatively smaller size of our modified U-Net (Table 4).

As mentioned previously, the task of table localization can be viewed either as a table detection task or a table segmentation task. Consequently, we compared

Table 4: Comparison of models by number of parameters

Model	# parameters (M=million)
TableBank [13]	$17M$
YOLO-v3 [17]	$65M$
pix2pixHD [23]	$188M$
CascadeTabNet [14]	$83M$
Original U-Net [14]	$8M$
Modified U-Net (Ours)	**3.5M**

its performance in these two task settings by customized training of two popular object detection models and three semantic segmentation models.

Detection models: For our experiments, we used TableBank [13], an open-source table detector trained on 163,417 MS Word documents and 253,817 LaTeX documents. We fine-tuned TableBank directly using the MediTables training set. We also trained a YOLO-v3 [17] object detection model. Unlike TableBank training, we followed the protocol used for our proposed model (i.e. pre-training on existing document datasets) (Sec. 5.2).

Segmentation models: We trained pix2pixHD [23], a popular image pixel-level image translation model. Keeping the relatively small size of our dataset in mind, we trained a scaled down version. In addition, we trained CascadeTabNet [14], a state-of-the-art segmentation model developed specifically for table segmentation. As a postprocessing step, we performed morphological closing on the results from segmentation approaches for noise reduction and filling holes in the output masks.

As a preliminary experiment, we examined performance when the models pre-trained on existing document datasets were directly evaluated on the MediTables test set (i.e. without any fine-tuning). As Table 3a shows, recent models developed specifically for table localization (CascadeTabNet [14], TableBank [13]) show good performance. Our modified U-Net's performance is relatively inferior, likely due to its inability to bridge the domain gap between existing datasets and medical domain documents directly.

Upon fine-tuning the document dataset pre-trained models data from the domain (i.e.MediTables dataset), a very different picture emerges (Table 3b). Our proposed approach (modified U-Net) outperforms strong baselines and existing approaches, including the models customized for table localization. We hypothesize that this is due to the ability of our modified U-Net to judiciously utilize the within-domain training data to close the gap between pre-trained setting and fine-tuned setting in an effective manner. Another important observation is that the deviation from average in our model is typically the smallest compared to other models. Finally, it is important to note that superior performance has been achieved by our model even though it is drastically smaller in size compared to customized and state of the art baselines (see Table 4).

Fig. 3: Table localization results for various models for images with highest IoU score using our model's predictions.

Performance comparison by table type: The previous experiments focused on evaluation for all tables. To examine performance by table type (T1, T2), the previous 2-class (table, background) formulation was replaced by a 3-class prediction setup (T1, T2 and background). As Table 5 shows, our modified U-Net once again performs the best across table types and measures.

Figure 3 shows qualitative results on images with the highest IoU score as per our model's predictions. The superior quality of our results is evident. A similar set of results can be viewed by table type in Figure 4. Images with the lowest IoU score as per our model prediction can be viewed in Figure 5. These represent the most challenging images. As mentioned previously, a high degree of skew and small footprint of the table in the image generally affect the performance

Table 5: Per-table type performance of models pre-trained on existing document datasets and fine-tuned on MediTables (PPA is not defined for detection models)

Model	T1			T2		
	IOU (%)	PPA (%)	F1 (%)	IOU (%)	PPA (%)	F1 (%)
TableBank [13]	92.59 ± 11.27	NA	95.75 ± 00.07	84.90 ± 16.20	NA	90.95 ± 10.30
YOLO-v3 [17]	30.88 ± 02.83	NA	47.18 ± 14.47	56.38 ± 07.58	NA	72.10 ± 06.19
pix2pixHD [23]	85.42 ± 01.65	92.74 ± 06.82	92.13 ± 06.38	92.67 ± 01.34	99.10 ± 00.89	96.19 ± 00.71
CascadeTabNet [14]	92.66 ± 10.87	93.22 ± 10.37	95.84 ± 06.39	93.44 ± 10.21	93.76 ± 10.22	96.29 ± 06.14
Modified U-Net (Ours)	$\mathbf{95.48 \pm 03.82}$	$\mathbf{99.08 \pm 00.91}$	$\mathbf{97.69 \pm 01.96}$	$\mathbf{94.30 \pm 00.30}$	$\mathbf{99.62 \pm 00.37}$	$\mathbf{97.06 \pm 01.61}$

from our model. However, even for these images, the quality of results is quite acceptable.

7 Conclusion

In this paper, we have presented a dataset for diverse healthcare and medical document images. We hope that our efforts encourage the community to expand our dataset and build upon our findings to enable richer understanding of medical document images. Given its distinct nature, we also expect our dataset to be considered along with existing datasets when benchmarking new table localization approaches in future.

We have also proposed a compact yet high performing approach for localizing and categorizing tables in medical documents. The performance of the proposed approach is greatly facilitated by our choice of using a segmentation network (as opposed to a detection network), the skip-connectivity for enhanced gradient flow and by our choice of losses and training procedure. Our model has the potential to operate as the first step in a processing pipeline for understanding tabular content in medical documents. Another significant advantage is the compact size of our model, making it potentially attractive for deployment on mobile and embedded devices.

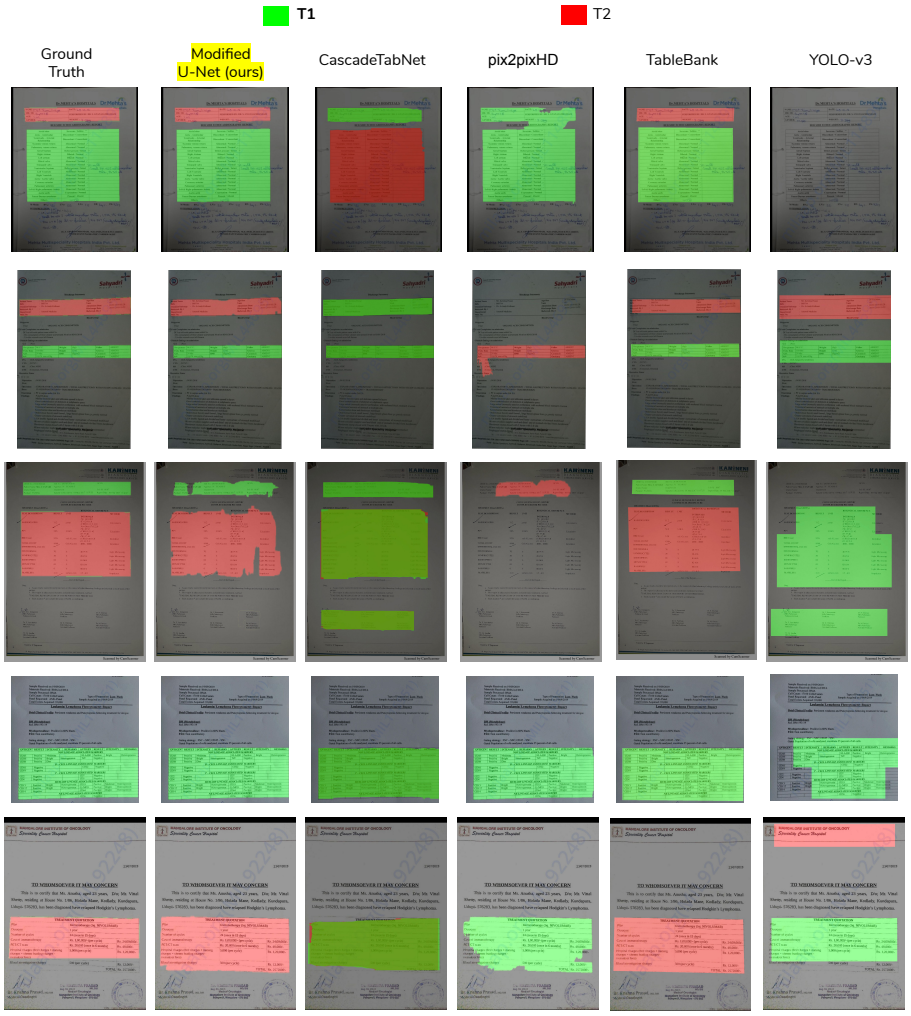

Fig. 4: Tables T1 (green) and T2 (red) segmentation results of three semantic segmentation models and two object detection models on the testing set of MediTables. Note that predictions (colors) may also be incorrect in terms of table type labels (T1,T2) in some instances.

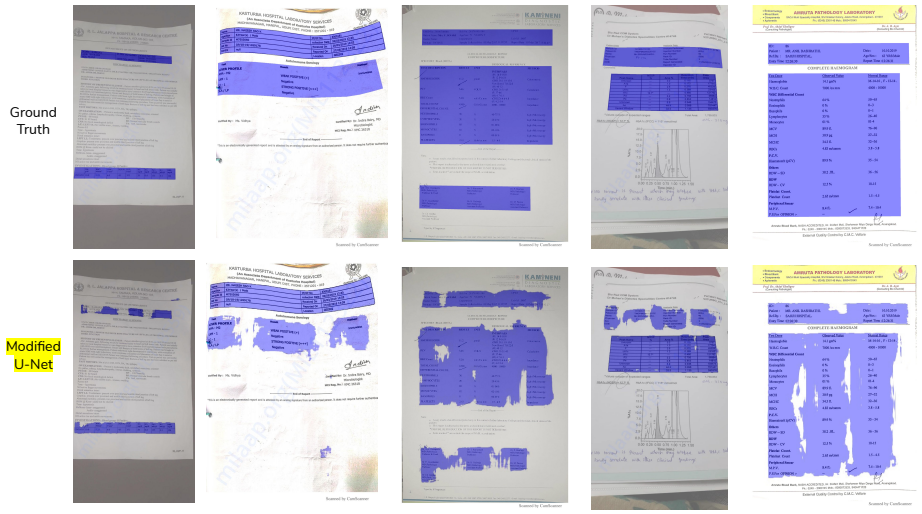

Fig. 5: Table localization results for various models for images with lowest IoU score using our model's predictions.

References

1. Fang, J., Gao, L., Bai, K., Qiu, R., Tao, X., Tang, Z.: A table detection method for multipage pdf documents via visual seperators and tabular structures. In: 2011 International Conference on Document Analysis and Recognition. pp. 779–783. IEEE (2011) 3
2. Gao, L., Yi, X., Jiang, Z., Hao, L., Tang, Z.: Icdar2017 competition on page object detection. In: 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). vol. 1, pp. 1417–1422. IEEE (2017) 6
3. Gilani, A., Qasim, S.R., Malik, I., Shafait, F.: Table detection using deep learning. In: 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). vol. 1, pp. 771–776. IEEE (2017) 3
4. Girshick, R.: Fast r-cnn. In: Proceedings of the IEEE international conference on computer vision. pp. 1440–1448 (2015) 3
5. Göbel, M., Hassan, T., Oro, E., Orsi, G.: Icdar 2013 table competition. In: 2013 12th International Conference on Document Analysis and Recognition. pp. 1449–1453. IEEE (2013) 6
6. Goyal, M., Yap, M.H., Hassanpour, S.: Multi-class semantic segmentation of skin lesions via fully convolutional networks. arXiv preprint arXiv:1711.10449 (2017) 7
7. Ha, J., Haralick, R.M., Phillips, I.T.: Recursive xy cut using bounding boxes of connected components. In: Proceedings of 3rd International Conference on Document Analysis and Recognition. vol. 2, pp. 952–955. IEEE (1995) 3
8. Hao, L., Gao, L., Yi, X., Tang, Z.: A table detection method for pdf documents based on convolutional neural networks. In: 2016 12th IAPR Workshop on Document Analysis Systems (DAS). pp. 287–292. IEEE (2016) 3
9. Huang, Y., Yan, Q., Li, Y., Chen, Y., Wang, X., Gao, L., Tang, Z.: A yolo-based table detection method. In: 2019 International Conference on Document Analysis and Recognition (ICDAR). pp. 813–818. IEEE (2019) 3
10. Kavasidis, I., Pino, C., Palazzo, S., Rundo, F., Giordano, D., Messina, P., Spampinato, C.: A saliency-based convolutional neural network for table and chart detection in digitized documents. In: International Conference on Image Analysis and Processing. pp. 292–302. Springer (2019) 3
11. Kieninger, T., Dengel, A.: The t-recs table recognition and analysis system. In: International Workshop on Document Analysis Systems. pp. 255–270. Springer (1998) 3
12. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. nature **521**(7553), 436 (2015) 2
13. Li, M., Cui, L., Huang, S., Wei, F., Zhou, M., Li, Z.: Tablebank: Table benchmark for image-based table detection and recognition. arXiv preprint arXiv:1903.01949 (2019) 6, 7, 8, 10
14. Prasad, D., Gadpal, A., Kapadni, K., Visave, M., Sultanpure, K.: Cascadetabnet: An approach for end to end table detection and structure recognition from image-based documents (2020) 3, 7, 8, 10
15. Prusty, A., Aitha, S., Trivedi, A., Sarvadevabhatla, R.K.: Indiscapes: Instance segmentation networks for layout parsing of historical indic manuscripts. In: 2019 International Conference on Document Analysis and Recognition (ICDAR). pp. 999–1006. IEEE (2019) 7
16. Rahman, M.A., Wang, Y.: Optimizing intersection-over-union in deep neural networks for image segmentation. In: International symposium on visual computing. pp. 234–244. Springer (2016) 6

17. Redmon, J., Farhadi, A.: Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767 (2018) 7, 8, 10
18. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015) 5
19. Sarvadevabhatla, R.K., Dwivedi, I., Biswas, A., Manocha, S.: Sketchparse: Towards rich descriptions for poorly drawn sketches using multi-task hierarchical deep networks. In: Proceedings of the 25th ACM international conference on Multimedia. pp. 10–18 (2017) 7
20. Schreiber, S., Agne, S., Wolf, I., Dengel, A., Ahmed, S.: Deepdesrt: Deep learning for detection and structure recognition of tables in document images. In: 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). vol. 1, pp. 1162–1167. IEEE (2017) 3
21. Siddiqui, S.A., Malik, M.I., Agne, S., Dengel, A., Ahmed, S.: Decnt: Deep deformable cnn for table detection. IEEE Access **6**, 74151–74161 (2018) 3
22. Taghva, K., Nartker, T., Borsack, J., Condit, A.: Unlv-isri document collection for research in ocr and information retrieval **3967** (02 2000) 6
23. Wang, T.C., Liu, M.Y., Zhu, J.Y., Tao, A., Kautz, J., Catanzaro, B.: pix2pixhd: High-resolution image synthesis and semantic manipulation with conditional gans 7, 8, 10
24. Yang, X., Yumer, E., Asente, P., Kraley, M., Kifer, D., Lee Giles, C.: Learning to extract semantic structure from documents using multimodal fully convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 5315–5324 (2017) 3
25. Yildiz, B., Kaiser, K., Miksch, S.: pdf2table: A method to extract table information from pdf files. In: IICAI. pp. 1773–1785 (2005) 3