SLANet-1M: A Lightweight and Efficient Model for Table Recognition with Minimal Computational Cost

Nguinwa Mbakop Dimitri Romaric
University of Florence
iCoSys - HEIA-FR, HES-SO
Fribourg, Switzerland
nguinwa.dimitri@edu.unifi.it

Andrea Petrucci
iCoSys - HEIA-FR, HES-SO
Fribourg, Switzerland
andrea.petrucci@hefr.ch

Jean Hennebert
iCoSys - HEIA-FR, HES-SO
Fribourg, Switzerland
jean.hennebert@hefr.ch

Simone Marinai
DINFO
University of Florence
Florence, Italy
simone.marinai@unifi.it

Abstract

Modern approaches for table recognition consist of an encoder for feature extraction and one or more decoders for structure recognition and cell box detection. Recent advancements in this field have introduced Transformers, initially in the decoders and more recently in the encoder as well. While these improvements have enhanced performance, they have also increased model complexity, requiring larger datasets for training, a pre-training step, and higher inference time.

In this paper, we explore SLANet, a lightweight transformer-free model originally trained on PubTabNet. To train a more robust version, we combined two publicly available datasets (PubTabNet and SynthTabNet) into one dataset of 1 million of images table, which led us to name the resulting model **SLANet-1M**. On PubTabNet, SLANet-1M improves the original SLANet's S-TEDS score by **0.35%**. It also scores only **0.53%** below the state-of-the-art UniTable Large, while using nearly **14 times fewer parameters**. SLANet*—a variant trained on PubTabNet and a quarter of SynthTabNet— achieves a 0.47% improvement. On SynthTabNet, SLANet-1M performs exceptionally well, with an S-TEDS score just **0.03%** lower than UniTable Large. Additionally, SLANet-1M outperforms major large vision-language models (VLMs) like GPT-4o, Granite Vision, and Llama Vision on this specific table recognition task. **SLANet-1M is also more efficient during inference**, offering faster processing and CPU-friendly execution, eliminating the need for a GPU.

1 Introduction

Tables contain a wealth of information in a concise format and are prevalent in documents. Extracting table information accurately is crucial for many applications (data analysis, finance, health, and so on). The table recognition task focuses on detecting tables in image-based documents and extracting their structure and contents in HTML format. However, due to the complexity of tables—such as rowspan, colspan, and multi-header layouts—table recognition remains a challenging task, even for advanced large vision-language models (VLMs) like GPT-4o (OpenAI, 2024), GPT-4-turbo (Yang et al., 2023), Granite Vision 3.2 (Team et al., 2025), and Llama Vision 3.2 (AI, 2025).

This paper presents a solution for companies that require high-performance table recognition without extensive computational resources. We enhanced SLANet (Li et al., 2022) quantitatively and qualitatively by training it on additional data, demonstrating that a lightweight model without Transformers can achieve performance comparable to more complex transformer-based models. Furthermore, we show that our improved SLANet is faster than state-of-the-art (SOTA) models while maintaining high accuracy.

We name this enhanced model **SLANet-1M** as it is trained on 1 million images by combining PubTabNet[1] and SynthTabNet[2] datasets.

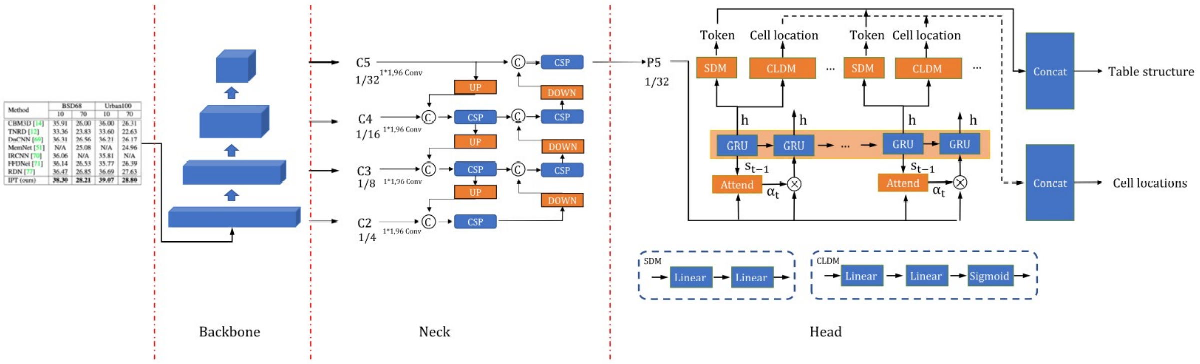
2 Related works

Many recent models on table recognition task have demonstrated great performance. Here we explore some of them, in particular models that follow the encoder-decoder architecture. We show that with the introduction of Transformers, their structure has adopted this technology firstly in the decoders and subsequently in the encoder as well.

2.1 EDD

The Encoder-Dual-Decoder (EDD) model was introduced in the PubTabNet paper (Zhong et al., 2020). EDD consists of an encoder, an attention-based structure decoder, and an attention-based cell decoder. The use of two decoders stems from the

[1] https://github.com/ibm-aur-nlp/PubTabNet
[2] https://github.com/IBM/SynthTabNet



Method	BSD68		Urban100	
	10	70	10	70
CBM3D [14]	35.91	26.00	36.00	26.31
TNRD [12]	33.36	23.83	33.60	22.63
DnCNN [69]	36.31	26.56	36.21	26.17
MemNet [51]	N/A	25.08	N/A	24.96
IRCNN [70]	36.06	N/A	35.81	N/A
FFDNet [71]	36.14	26.53	35.77	26.39
RDN [77]	36.47	26.85	36.69	27.63
IPT (ours)	**38.30**	**28.21**	**39.07**	**28.80**

Figure 1: Architecture of SLANet.

observation that table structure recognition and cell content recognition are distinct tasks that are inefficient to solve with a single attention-based decoder.

EDD's encoder is a convolutional neural network (CNN) that captures visual features from input table images. The structure decoder and the cell decoder are recurrent neural networks (RNNs) equipped with an attention mechanism to process and reconstruct the structure and content of the table.

2.2 Table Master

Table Master (Ye et al., 2021) was introduced as a solution for the ICDAR 2021 competition on scientific literature parsing (Task B: table recognition to HTML). Inspired by MASTER (Lu et al., 2021), its decoder is composed of Transformer decoder layers.

Table Master employs two decoder branches, each consisting of three Transformer decoder layers, with the first layer shared between both branches. One branch is responsible for predicting the HTML sequence, while the other conducts box regression. Unlike other models that split tasks at the final layer, Table Master decouples sequence prediction and box regression immediately after the first Transformer decoder layer.

2.3 TableFormer

Introduced in the SynthTabNet paper (Nassar et al., 2022), TableFormer employs an hybrid CNN-Transformer architecture as encoder. The encoder consists of a ResNet-18 CNN and a Transformer encoder with two encoder layers, extracting features from input images into a fixed-length feature vector. TableFormer has two decoders: a structure decoder, modeled as a Transformer decoder with four decoder layers, incorporating multi-head attention and feed-forward networks (FFNs), and a cell box decoder, which utilizes the same Transformer encoder and decoder but introduces an additional attention-based FFN block to refine cell-level predictions.

2.4 VAST

The Visual-Alignment Sequential Coordinate Table Recognizer (VAST) (Huang et al., 2023) consists of three primary components: a modified ResNet enhanced with multi-aspect global content attention as the CNN-based image encoder, a transformer-based HTML sequence decoder, and a Transformer block for coordinate sequence decoding, allowing precise localization of table structures.

2.5 UniTable

UniTable (Peng et al., 2024) is the most recent model in table recognition, introducing a transformer-based encoder alongside a Transformer decoder. Initially, in an earlier attempt (Huang et al., 2023), replacing the CNN encoder with a vanilla Transformer with linear projection led to a performance drop compared to models using CNN or hybrid CNN-Transformer encoders.

To address this issue, UniTable implements self-supervised pre-training for the visual encoder.

2.6 SLANet

SLANet stands for Structure Location Alignment Network, presented in PP-StructureV2 (Li et al., 2022) as an efficient Table Recognition algorithm. In Figure 1 we show the network architecture of the model, composed of a backbone, a neck, and a head. we provide a detailed description of the architecture in Section 4.2.

3 Contribution

Our main contribution lies in adapting and evaluating the SLANet model (Li et al., 2022) on an additional dataset to assess its generalization capabilities and performance relative to state-of-the-art (SOTA) methods. Detailed information on the implementation and training procedure is provided in Sections 6.1.1 and 6.1.2.

In addition, we extend prior work by evaluating and comparing the **inference time on CPU** of some of the models discussed in the previous section—an aspect that has not been systematically analyzed in their original studies.

4 Formulation and SLANet's details

In this section, we define the table structure recognition task and provide a detailed description of the model we adopt for our experiments. We also outline the loss functions used during training.

4.1 Task Definition

The objective of **Table Recognition (TR)** is to convert a tabular image I into a structured, machine-readable format T, capturing both its *logical* and *physical* structure. The logical structure is often represented in HTML format, denoted as a tokenized sequence $S = [s_1, ..., s_T]$, where each s corresponds to an HTML tag. The physical structure consists of the bounding box coordinates of non-empty cells, represented as $B = [b_1, ..., b_N]$, where each bounding box is defined as $b = (x_{\text{min}}, y_{\text{min}}, x_{\text{max}}, y_{\text{max}})$, with integer values. Additionally, $C = [c_1, ..., c_N]$ represents the textual content inside each cell, following a reading order. While the number of elements in B and C are the same, they are typically fewer than those in S ($N < T$), since the HTML sequence includes both filled and empty cells. Each cell is associated with a single bounding box and may contain either a single line or multiple lines of text.

4.2 SLANet's Architecture

4.2.1 Backbone

SLANet employs PP-LCNet (Cui et al., 2021) as its backbone, a lightweight, CPU-friendly convolutional neural network architecture. PP-LCNet introduced several novel ideas to improve the accuracy without increasing the inference time. These techniques can be summarize as follows:

- Better activation function; from ReLU to H-Swish.
- SE (squeeze-and-excitation) modules (Hu et al., 2018) at appropriate positions.
- Larger convolution kernels; replacing the 3x3 convolutional kernels with the 5x5 convolutional kernels only at the tail of the network.
- Larger dimensional 1x1 conv layer after GAP; to give the network a stronger fitting ability and allow for more storage of the model with little increase of inference time. PP-LCNet appended a 1280-dimensional size 1x1 conv (equivalent to FC layer) after the final GAP layer.

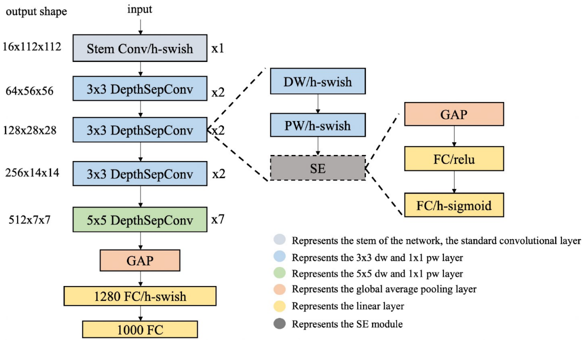


Figure 2: PP-LCNet. PP-LCNet includes optional modules, indicated by the dotted box. The stem section utilizes a standard 3×3 convolution. DepthSepConv refers to depth-wise separable convolutions, where DW stands for depth-wise convolution, PW denotes point-wise convolution, and GAP represents Global Average Pooling.

PP-LCNet uses DepthSepConv (Howard et al., 2017) as its basic block, the architecture is shown in Figure 2. Depthwise Separable Convolution is a good alternative to the classic convolution, as it can reduce the complexity and improve the inference speed of the operation while maintaining the accuracy. With all these improvements, PP-LCNet achieves better performance on multiple tasks with respect to lightweight models such as ShuffleNetV2 (Ma et al., 2018), MobileNetV3 (Howard et al., 2019), and GhostNet (Han et al., 2020).

4.2.2 Neck

SLANet enhances feature fusion to effectively address challenges caused by scale variations in complex scenes. To achieve this efficiently, it utilizes CSP-PAN (Yu et al., 2021), which integrates the PAN (Path Aggregation Network) structure for multi-level feature extraction and the CSP (Cross Stage Partial) structure for feature concatenation and fusion between adjacent feature maps.

Path Aggregation Network (PAN) (Liu et al., 2018) improves the feature pyramid by enhancing localization accuracy and optimizing information flow. It introduces:

- Bottom-up path augmentation, which shortens the information path and strengthens low-level features with precise localization signals.
- Adaptive feature pooling, which aggregates features across all levels for each proposal, ensuring a more structured and efficient feature propagation while avoiding arbitrary assignments.

These enhancements create more efficient and structured feature pathways, improving feature fusion and ultimately boosting detection performance.

Cross Stage Partial (CSP) Structure (Wang et al., 2020) is designed to enhance gradient flow while reducing computational cost. It achieves this by splitting the base layer's feature map into two parts and merging them through a cross-stage hierarchy. By dividing the gradient flow into separate network paths, CSP ensures that the propagated gradient information exhibits a greater correlation difference, improving learning efficiency through alternating concatenation and transition steps.

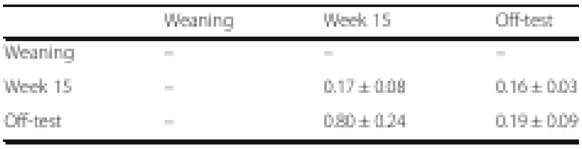
To optimize efficiency further, SLANet reduces the output channels of CSP-PAN from 128 to 96, effectively decreasing the model size without compromising performance.

4.2.3 Head

In its head module, SLANet employs a GRU along with two key components: the **Structure Decode Module** (SDM) and the **Cell Location Decode Module** (CLDM). The result of the feature fusion is passed in the GRU, and at each step, the GRU's output is concatenated and passed to both SDM and CLDM, generating cell tokens and their corresponding bounding box coordinates.

SLANet ensures one-to-one alignment between cell tokens and their coordinates, with SLAHead responsible for maintaining this correspondence. The tokens and coordinates from all decoding steps are concatenated to construct the HTML table representation along with the precise coordinates of all cells.

Inspired by TableMaster (Ye et al., 2021), SLANet treats <td> and </td> as a single token (<td></td>), simplifying the tokenization process for table structure generation.



	Weaning	Week 15	Off-test
Weaning	–	–	–
Week 15	–	0.17 ± 0.08	0.16 ± 0.03
Off-test	–	0.80 ± 0.24	0.19 ± 0.09

Figure 3: An example image from PubTabNet.

4.3 Loss Functions

The total loss function consists of two components: *structure loss* and *localization loss*, combined as:

$$\mathcal{L}_{\text{total}} = \lambda_{\text{structure}}\mathcal{L}_{\text{structure}} + \lambda_{\text{loc}}\mathcal{L}_{\text{loc}}$$

This combined loss ensures the model effectively learns both table structure and bounding box localization.

4.3.1 Structure Loss

The structure loss measures the accuracy of table structure predictions using the cross-entropy loss:

$$\mathcal{L}_{\text{structure}} = -\frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{T} y_{i,j}\log(\hat{y}_{i,j})$$

where K is the batch size, T is the sequence length, $y_{i,j}$ is the ground truth token, and $\hat{y}_{i,j}$ is the predicted probability.

4.3.2 Localization Loss

The localization loss evaluates bounding box accuracy using the *SmoothL1* loss:

$$\textit{SmoothL1}(x) = \begin{cases} 0.5\,x^2, & \text{if } |x| < 1 \\ |x| - 0.5, & \text{otherwise} \end{cases}$$

where $x = \mathbf{b}_{i,j} - \hat{\mathbf{b}}_{i,j}$.
The localization loss is normalized as:

$$\mathcal{L}_{\text{loc}} = \frac{\sum_{i,j} \textit{SmoothL1}(\mathbf{b}_{i,j} - \hat{\mathbf{b}}_{i,j}) \cdot m_{i,j}}{\sum_{i,j} m_{i,j} + \epsilon}$$

where $\epsilon > 0$ prevents division by zero, $\mathbf{b}_{i,j} = (x_{\min}, y_{\min}, x_{\max}, y_{\max})$ is the ground truth bounding box, $\hat{\mathbf{b}}_{i,j}$ is the predicted bounding box, and $m_{i,j}$ is a mask for valid bounding boxes.

5 Datasets and Metrics

5.1 Datasets

In this paper, we explore two publicly available table structure recognition benchmark datasets: PubTabNet and SynthTabNet.

5.1.1 PubTabNet

The PubTabNet (Zhong et al., 2020) dataset consists of 509,892 annotated PNG images (500,777 for training and 9,115 for validation). Each table is annotated with its structure in HTML format, along with tokenized text and bounding boxes for each cell. As shown in Figure 3, the dataset primarily contains simpler table structures with relatively few rows and columns. Additionally, the dataset exhibits limited variation in table styles, which hinders model generalization to unseen table formats. Recognizing these limitations, the authors of TableFormer introduced SynthTabNet to address these issues.

5.1.2 SynthTabNet

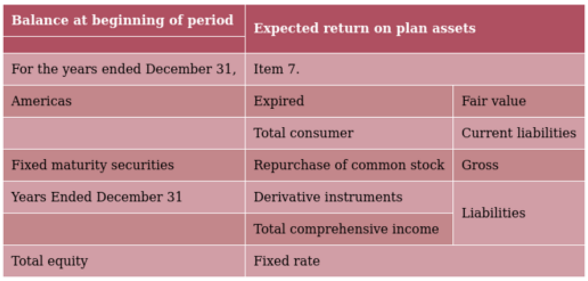
SynthTabNet (Nassar et al., 2022) is a large-scale synthetically generated dataset designed to offer control over dataset size, table structures, table styles, and content types.

The dataset aims to overcome the shortcomings of PubTabNet and FinTabNet, which suffer from skewed distributions toward simpler tables, limited stylistic diversity, and restricted cell content types. SynthTabNet consists of 600,000 tables, divided into four 150,000-table subsets:

Finance (1) and *PubTabNet* (3), which mimic FinTabNet[3] and PubTabNet while incorporating more complex structures. *Marketing* (2), which features high-contrast, colorful tables that resemble real-world marketing documents as shown in Figure 4. *Sparse* (4), which contains tables with minimal content, testing model performance on incomplete or sparsely populated tables. All parts are divided into Train, Val, and Test splits (80%, 10%, 10%). Because SynthTabNet provides a comprehensive evaluation of table recognition models across diverse table structures, we use it for ablation studies and present results separately for each subset.

5.2 PubTables-1M

Although we did not use PubTables-1M (Smock et al., 2022) in our experiments, we include it here as it is one of the largest table recognition (TR) datasets. PubTables-1M comprises nearly one million tables extracted from scientific articles, supports multiple input modalities, and provides detailed header and location information for table structures. These features make it a valuable resource for various modeling approaches.



Balance at beginning of period	Expected return on plan assets	
For the years ended December 31,	Item 7.	
Americas	Expired	Fair value
	Total consumer	Current liabilities
Fixed maturity securities	Repurchase of common stock	Gross
Years Ended December 31	Derivative instruments	Liabilities
	Total comprehensive income	
Total equity	Fixed rate	

Figure 4: An example image from SynthTabNet (Marketing subset).

However, as noted by UniTable (Peng et al., 2024), PubTables-1M suffers from several inconsistencies, particularly in its annotation method. The dataset uses word-wise bounding box (bbox) annotations, whereas PubTabNet and SynthTabNet follows a cell-wise annotation approach.

- Cell-wise annotation assigns a single bbox per table cell, allowing for a direct mapping between non-empty cells and their corresponding HTML structure.
- Word-wise annotation, used in PubTables-1M, assigns a bbox to each individual word, making it challenging to integrate with the table structure as effectively as cell-wise annotation.

This fundamental difference limits the general applicability of PubTables-1M for certain table recognition tasks.

5.3 Metrics

5.3.1 Accuracy

Used during the training the accuracy refers to the proportion of correctly identified table elements (such as structure, cells, or text) compared to the total number of ground truth elements. It measures the effectiveness of a table recognition system in correctly detecting and extracting tables from documents.

It is defined as:

$$\text{Acc.} = \frac{\text{Numb. of Correctly Recognized Elements}}{\text{Total Numb. of Ground Truth Elements}}$$

5.3.2 TEDS

TEDS (Tree-edit-distance-based Similarity), introduced by PubTabNet (Zhong et al., 2020), converts the table into a tree structure in HTML format

[3] https://developer.ibm.com/data/fintabnet/

and measures the edit distance between the prediction T_{pred} and the groundtruth T_{gt}. A shorter edit distance indicates a higher degree of similarity, leading to a higher TEDS score. TEDS measures both the table structure and table cell content. We also use S-TEDS as metric where only the table structure is considered. For comparison we consider more S-TEDS because for the content of cells some models rely on external text detection and text recognition models, which can differ from model to model and so can compromise the comparison.

TEDS between two trees is computed as:

$$TEDS = 1 - \frac{\textit{EditDist}(T_{gt}, T_{pred})}{\max(|T_{gt}|, |T_{pred}|)} \tag{1}$$

where *EditDist* denotes tree-edit distance (Pawlik and Augsten, 2016), and $|T|$ is the number of nodes in T.

Datasets	Records		Size (GB)	
Name	Train	Val	Train	Val
PubTabNet	500,777	9,115	11.6	0.2
SynthTabNet	480,347	59,618	24.2	3.0
Merged	**981,124**	**68,733**	**35.8**	**3.2**

Table 1: Dataset details including records and sizes for training and validation for SLANet-1M.

Models	Datasets	
	PubTabNet	SynthTabNet 3
SLANet	76.35	17.21
SLANet*	**77.07**	**81.72**

Table 2: Results (accuracy) of the first experiment, SLANet is the original model trained on PubTabNet, and SLANet* is the model trained on both PubTabNet and SynthTabNet part 3.

Models	Datasets			
	PubTabNet		SynthTabNet 3	
	TEDS	S-TEDS	TEDS	S-TEDS
SLANet	**95.89**	97.01	89.01	95.65
SLANet*	95.83	**97.48**	**92.87**	**99.47**

Table 3: Results (TEDS and S-TEDS) of the first experiment, SLANet is the original model trained on PubTabNet, and SLANet* is the model trained on both PubTabNet and SynthTabNet part 3.

6 Experiments and Results

6.1 Experiments

6.1.1 Implementations

We conducted two setup experiments, both on a 48G A40 GPU device, during 50 epochs using Adam as optimizer, the initial learning rate is set to 0.001 and adjusted to 0.0001 and 0.00005 after 29 and 39 epochs. The batch size is set to 48 for the first experiment and to 72 for the second.

6.1.2 Training

For the first experiment we trained SLANet from scratch on the PubTabNet and the third part of SynThTabNet for a total of 620,772 images for the training set, validate on the validation set of PubTabNet and tested on the same set because there is no the groundthruth fot the test set of PubTabNet.

For the second experiment we merged both datasets (PubTabNet and SynthTabNet) as detailed in Table 1. The validation set is obtained by merging all the validation sets of subsets of SynthTabNet with the validation set of PubTabNet. The tests are made on the test sets of SynthTabNet subsets.

6.2 Results

6.2.1 First Experiment

The model obtained with the first training setup is named **SLANet*** and the Table 2 and Table 3 summarize the performance of SLANet and SLANet* across PubTabNet and SynthTabNet (Part 3). SLANet*, trained on both datasets, consistently outperforms the original SLANet. On PubTabNet, SLANet* achieves a slight **0.72%** improvement in accuracy while maintaining comparable TEDS performance and a **0.47%** increase in S-TEDS.

The performance boost is more pronounced on SynthTabNet (Part 3), where SLANet* significantly surpasses SLANet, improving accuracy from 17.21% to 81.72%. Additionally, it demonstrates a substantial increase in TEDS (**+3.86%**) and S-TEDS (**+3.82%**), confirming its enhanced adaptability when trained on a more diverse dataset.

Table 4 compares SLANet* to state-of-the-art models on PubTabNet. Despite having **significantly fewer parameters** —nearly 14 times fewer than the strongest models — SLANet* achieves competitive performance. It is only **0.41%** on S-TEDS score behind the SOTA UniTable Large, demonstrating its efficiency and effectiveness in the table recognition task.

Models	**TEDS**	**S-TEDS**	**SIZE (M)**
EDD (Zhong et al., 2020)	88.30	89.90	-
TableMaster (Ye et al., 2021)	96.12	97.56	253
TableFormer (Nassar et al., 2022)	93.60	96.75	53.2
VAST (Huang et al., 2023)	96.31	97.23	-
UniTable Base (Peng et al., 2024)	94.78	95.63	30
UniTable Large (Peng et al., 2024)	**96.50**	**97.89**	125
SLANet (Li et al., 2022)	95.89	97.01	**9.2**
SLANet* (ours)	95.83	97.48	**9.2**

Table 4: Comparison on PubTabNet of models based on TEDS, S-TEDS, and SIZE.

Models	**S-TEDS**	**Size (M)**
TableFormer	96.70	53.2
UniTable Base	98.97	30
UniTable Large	**99.39**	125
SLANet-1M	99.36	**9.2**

Table 5: Comparison of performance on SynthTabNet.

6.2.2 Second Experiment

In the second experiment, we trained SLANet on the consolidated dataset detailed in Table 1. The resulting model, referred to as SLANet-1M, demonstrates strong performance on the SynthTabNet benchmark, as illustrated in Table 5. In particular, SLANet-1M lags behind UniTable Large by a mere **0.03%**, despite possessing approximately 14 times fewer parameters. It is important to highlight that UniTable Large benefits from a significantly broader training regimen—having been trained on PubTabNet, SynthTabNet, and FinTabNet for table recognition, in addition to undergoing a pre-training phase on PubTabNet, SynthTabNet, FinTabNet, and PubTables-1M.

6.2.3 Ablation Study

Table 6 presents an ablation study comparing the S-TEDS scores of UniTable, SLANet, SLANet*, and SLANet-1M across the four subsets of the SynthTabNet dataset. As expected, SLANet-1M outperforms both SLANet and SLANet* on all the three other subsets, given that it was explicitly trained on these data partitions. Notably, SLANet-1M also demonstrates a modest improvement of **0.05%** on the PubTabNet subset of SynthTabNet.

When compared to UniTable Large, SLANet-1M achieves superior performance on the Marketing subset with a **0.14%** lead and matches UniTable's score on the Sparse subset. On the PubTabNet subset, it trails slightly by only **0.04%**. The most pronounced difference is observed on the Finance subset, where SLANet-1M falls behind UniTable Large by **0.23%**—this being the only subset where UniTable Base also surpasses SLANet-1M, albeit by a smaller margin of **0.06%**. This performance gap can likely be attributed to UniTable's broader training scope, as it was trained on a more diverse set of datasets, including FinTabNet, which may contribute to its enhanced generalization on financial tables.

7 Qualitative Results and Inference Time

7.1 Qualitative Results

In this section, we present a qualitative analysis by first comparing SLANet-1M with the original SLANet, followed by a comparison with several large vision-language models (VLMs). One representative sample per configuration was retained. Additional examples can be found in the appendix.

7.1.1 SLANet vs SLANet-1M

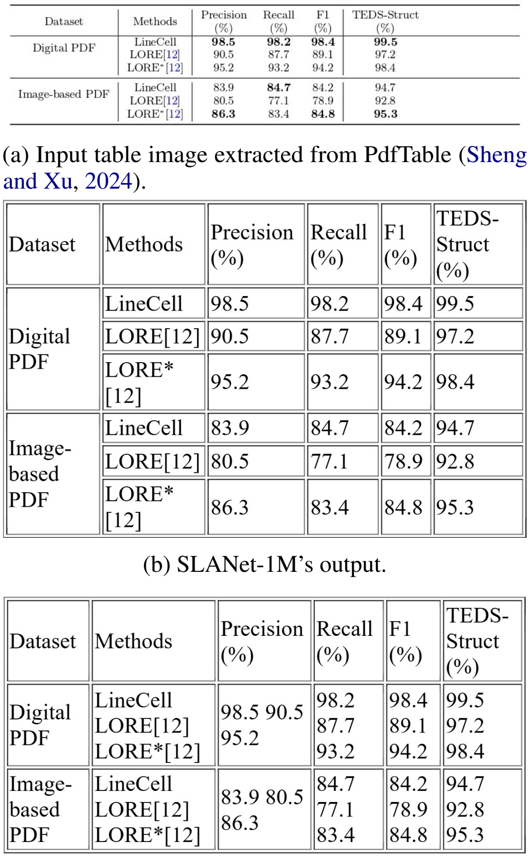
Figure 5 illustrates the inputs provided to both SLANet and SLANet-1M, along with the corresponding HTML tables generated by each model. As shown, SLANet encounters difficulties in accurately identifying and separating the correct number of rows. In contrast, SLANet-1M successfully overcomes this limitation, generating a well-structured HTML table that clearly delineates rows, even in cases where they are not explicitly wired in the input.

7.1.2 SLANet vs VLMs

Following the approach of UniTable (Peng et al., 2024), we conduct a qualitative comparison between our model and several state-of-the-art large vision-language models (VLMs). Figure 6 presents the input image alongside the outputs generated by SLANet-1M, GPT-4o (OpenAI, 2024), Granite

Models	**Finance**	**Marketing**	**PubTabNet**	**Sparse**
UniTable Base (Peng et al., 2024)	99.41	98.35	99.44	98.69
UniTable Large (Peng et al., 2024)	**99.58**	99.08	**99.56**	99.34
SLANet (Li et al., 2022)	89.83	80.83	95.65	86.10
SLANet* (ours)	91.26	82.99	99.47	91.33
SLANet-1M (ours)	99.35	**99.22**	99.52	**99.34**

Table 6: Comparison across different subsets of SynthTabNet dataset.



Dataset	Methods	Precision (%)	Recall (%)	F1 (%)	TEDS-Struct (%)
Digital PDF	LineCell	**98.5**	**98.2**	**98.4**	**99.5**
	LORE[12]	90.5	87.7	89.1	97.2
	LORE*[12]	95.2	93.2	94.2	98.4
Image-based PDF	LineCell	83.9	**84.7**	84.2	94.7
	LORE[12]	80.5	77.1	78.9	92.8
	LORE*[12]	**86.3**	83.4	**84.8**	**95.3**

(a) Input table image extracted from PdfTable (Sheng and Xu, 2024).

Dataset	Methods	Precision (%)	Recall (%)	F1 (%)	TEDS-Struct (%)
Digital PDF	LineCell	98.5	98.2	98.4	99.5
	LORE[12]	90.5	87.7	89.1	97.2
	LORE* [12]	95.2	93.2	94.2	98.4
Image-based PDF	LineCell	83.9	84.7	84.2	94.7
	LORE[12]	80.5	77.1	78.9	92.8
	LORE* [12]	86.3	83.4	84.8	95.3

(b) SLANet-1M's output.

Dataset	Methods	Precision (%)	Recall (%)	F1 (%)	TEDS-Struct (%)
Digital PDF	LineCell LORE[12] LORE*[12]	98.5 90.5 95.2	98.2 87.7 93.2	98.4 89.1 94.2	99.5 97.2 98.4
Image-based PDF	LineCell LORE[12] LORE*[12]	83.9 80.5 86.3	84.7 77.1 83.4	84.2 78.9 84.8	94.7 92.8 95.3

(c) SLANet's output.

Figure 5: Qualitative comparison between SLANet and SLANet-1M.

Vision 3.2 (Team et al., 2025), and Llama Vision 3.2 (AI, 2025).

We adopt the same prompt used in UniTable (Peng et al., 2024) and in the evaluation of the Optical Character Recognition (OCR) capabilities of GPT-4V (Shi et al., 2023): *"Please read the table in this image and return an HTML-style reconstructed table in text. Do not omit anything."*

The results show that SLANet-1M outperforms GPT-4o, which fails to preserve the correct number of rows and introduces unnecessary blank spaces and empty cells. In contrast, SLANet-1M more faithfully maintains the table's structural integrity. Among the baseline VLMs, Granite Vision 3.2 performs the best, although it misplaces the content of the first cell by rendering it in the last cell of the first row. Llama Vision 3.2 simplifies the output by reducing the table to just two columns, revealing its limitations in handling complex table structures.

One qualitative result is shown here; more quantitative and qualitative results are in Appendices A and B, respectively.

7.2 Inference Time

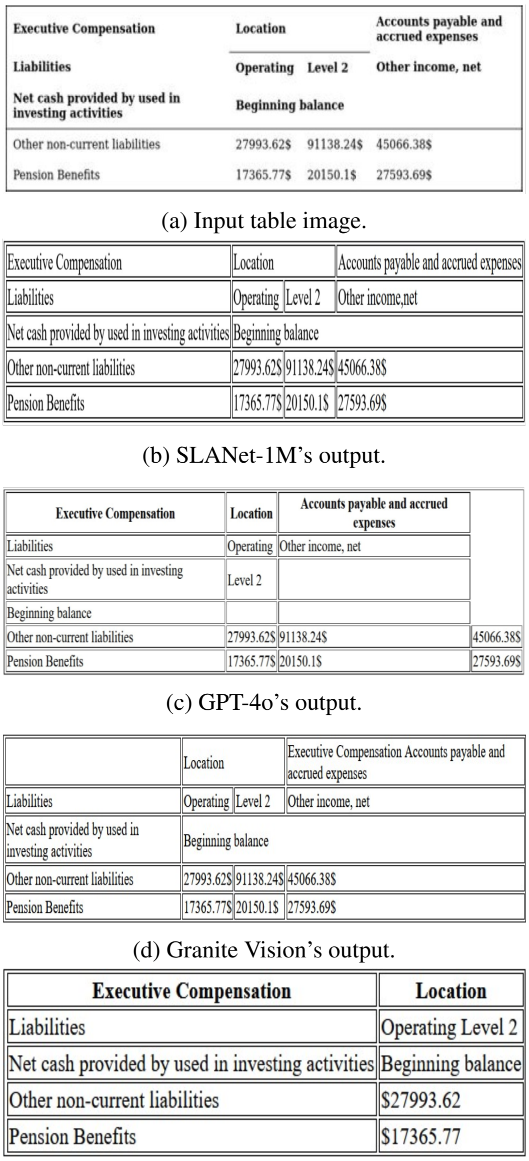
One of the main objectives of this research was to provide an alternative to transformer-based table recognition models—one that achieves similar performance while remaining efficient enough to run on a CPU with a satisfactory inference time.

All the models cited in this paper overlook this aspect. To address this, we compared the inference time of SLANet-1M (which is essentially the same as SLANet) against two state-of-the-art models: TableFormer and UniTable Large. The evaluation was conducted on a CPU-powered system with the following specifications:

- Processor: 11th Gen Intel(R) Core(TM) i7-11850H @ 2.50GHz, 2496 MHz, 8 Cores, 16 Logical Processors.
- Memory: 32.0 GB RAM.
- System Type: x64-based PC.
- Dataset: 200 images (50 images per subset).

The Docling technical report (Auer et al., 2024) highlights that TableFormer suffers from high inference time on CPU due to its reliance on EasyOCR[4], a finding that our experiments confirmed. Specifically, TableFormer exhibited an average inference time of 10,020 milliseconds, while UniTable Large was even slower, likely due to its fully transformer-based architecture, with an average inference time of 118,729 milliseconds. In contrast, SLANet-1M

[4] https://github.com/JaidedAI/EasyOCR



Executive Compensation	Location		Accounts payable and accrued expenses
Liabilities	Operating	Level 2	Other income, net
Net cash provided by used in investing activities	Beginning balance		
Other non-current liabilities	27993.62$	91138.24$	45066.38$
Pension Benefits	17365.77$	20150.1$	27593.69$

(a) Input table image.

Executive Compensation	Location		Accounts payable and accrued expenses
Liabilities	Operating	Level 2	Other income,net
Net cash provided by used in investing activities	Beginning balance		
Other non-current liabilities	27993.62$	91138.24$	45066.38$
Pension Benefits	17365.77$	20150.1$	27593.69$

(b) SLANet-1M's output.

Executive Compensation	Location	Accounts payable and accrued expenses	
Liabilities	Operating	Other income, net	
Net cash provided by used in investing activities	Level 2		
Beginning balance			
Other non-current liabilities	27993.62$	91138.24$	45066.38$
Pension Benefits	17365.77$	20150.1$	27593.69$

(c) GPT-4o's output.

	Location		Executive Compensation Accounts payable and accrued expenses
Liabilities	Operating	Level 2	Other income, net
Net cash provided by used in investing activities	Beginning balance		
Other non-current liabilities	27993.62$	91138.24$	45066.38$
Pension Benefits	17365.77$	20150.1$	27593.69$

(d) Granite Vision's output.

Executive Compensation	Location
Liabilities	Operating Level 2
Net cash provided by used in investing activities	Beginning balance
Other non-current liabilities	$27993.62
Pension Benefits	$17365.77

(e) Llama Vision's output.

Figure 6: Qualitative comparison between GPT-4o, Granite Vision, Llama Vision and SLANet-1M.

significantly outperformed both models, achieving an average inference time of less than **500 milliseconds**. The inference time refers to the time required to process the table, generate the HTML code, and save the result in Excel or CSV format.

8 Conclusion

In this paper, we evaluate SLANet on a new dataset and introduce SLANet-1M, a model trained on one million table images. We demonstrate both quantitatively and qualitatively that SLANet-1M outperforms SLANet and competes effectively with transformer-based architectures, and VLMs.

When trained on PubTabNet and the third subset of SynthTabNet, SLANet* achieves an S-TEDS score on PubTabNet that is only **0.41%** lower than the state-of-the-art (SOTA), despite using 14 times fewer parameters. When trained on PubTabNet and all subsets of SynthTabNet, its S-TEDS score on SynthTabNet is just **0.03%** below SOTA, maintaining the same efficiency.

Models	Inf. Time (ms)	Size (M)
TableFormer	10,020	53.2
UniTable Large	118,729	125
SLANet-1M	**463**	**9.2**

Table 7: Comparison of inference time on CPU.

Additionally, SLANet-1M offers faster inference time while being CPU-friendly, with only 9.2 million parameters. This makes it an ideal solution for users seeking a high-performance model without significant computational demands. Finally, we deployed SLANet-1M in the core engine of the Swiss AI center, making it accessible for those interested in testing it, it can be accessed here.

Limitations

Despite its many strengths, SLANet-1M does exhibit certain limitations. The most prominent among these is its dependence on external models for text detection and recognition. Additionally, due to its use of lightweight components, the quality of its predicted bounding boxes falls short compared to some state-of-the-art models in table recognition. Furthermore, since the majority of the training data comprises wireless tables, SLANet-1M encounters minor challenges in accurately interpreting the structure of fully wired tables. Notably, the latter limitation could be effectively mitigated through training on a more diverse and representative dataset.

Acknowledgements

This work was carried out within the framework of the Swiss-European Mobility Programme (SEMP), during a semester exchange from the University of Florence to the College of Engineering and Architecture of Fribourg. This work was also supported financially by the research grant Swiss AI Center 135788/IA-RECHERCHE24-11 and the institute iCoSys from the Swiss Universities of Applied Science HES-SO. We also thank the anonymous reviewers for their valuable feedback during the preparation of this paper.

References

Meta AI. 2025. Llama 3.2-vision: Large language and vision model. https://ollama.com/library/llama3.2-vision. Accessed: 2025-04-12.

Christoph Auer, Maksym Lysak, Ahmed Nassar, Michele Dolfi, Nikolaos Livathinos, Panos Vagenas, Cesar Berrospi Ramis, Matteo Omenetti, Fabian Lindlbauer, Kasper Dinkla, et al. 2024. Docling technical report. *arXiv preprint arXiv:2408.09869*.

Cheng Cui, Tingquan Gao, Shengyu Wei, Yuning Du, Ruoyu Guo, Shuilong Dong, Bin Lu, Ying Zhou, Xueying Lv, Qiwen Liu, et al. 2021. Pp-lcnet: A lightweight cpu convolutional neural network. *arXiv preprint arXiv:2109.15099*.

Kai Han, Yunhe Wang, Qi Tian, Jianyuan Guo, Chunjing Xu, and Chang Xu. 2020. Ghostnet: More features from cheap operations. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1580–1589.

Andrew Howard, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang, Yukun Zhu, Ruoming Pang, Vijay Vasudevan, et al. 2019. Searching for mobilenetv3. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1314–1324.

Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. 2017. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.

Jie Hu, Li Shen, and Gang Sun. 2018. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141.

Yongshuai Huang, Ning Lu, Dapeng Chen, Yibo Li, Zecheng Xie, Shenggao Zhu, Liangcai Gao, and Wei Peng. 2023. Improving table structure recognition with visual-alignment sequential coordinate modeling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11134–11143.

Chenxia Li, Ruoyu Guo, Jun Zhou, Mengtao An, Yuning Du, Lingfeng Zhu, Yi Liu, Xiaoguang Hu, and Dianhai Yu. 2022. Pp-structurev2: A stronger document analysis system. *arXiv preprint arXiv:2210.05391*.

Shu Liu, Lu Qi, Haifang Qin, Jianping Shi, and Jiaya Jia. 2018. Path aggregation network for instance segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8759–8768.

Ning Lu, Wenwen Yu, Xianbiao Qi, Yihao Chen, Ping Gong, Rong Xiao, and Xiang Bai. 2021. Master: Multi-aspect non-local network for scene text recognition. *Pattern Recognition*, 117:107980.

Ningning Ma, Xiangyu Zhang, Hai-Tao Zheng, and Jian Sun. 2018. Shufflenet v2: Practical guidelines for efficient cnn architecture design. In *Proceedings of the European conference on computer vision (ECCV)*, pages 116–131.

Ahmed Nassar, Nikolaos Livathinos, Maksym Lysak, and Peter Staar. 2022. Tableformer: Table structure understanding with transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4614–4623.

OpenAI. 2024. Gpt-4o technical report. https://openai.com/index/hello-gpt-4o/. Accessed: 2025-04-12.

Mateusz Pawlik and Nikolaus Augsten. 2016. Tree edit distance: Robust and memory-efficient. *Information Systems*, 56:157–173.

ShengYun Peng, Aishwarya Chakravarthy, Seongmin Lee, Xiaojing Wang, Rajarajeswari Balasubramaniyan, and Duen Horng Chau. 2024. Unitable: Towards a unified framework for table recognition via self-supervised pretraining. *arXiv preprint arXiv:2403.04822*.

Lei Sheng and Shuai-Shuai Xu. 2024. Pdftable: A unified toolkit for deep learning-based table extraction. *arXiv preprint arXiv:2409.05125*.

Yongxin Shi, Dezhi Peng, Wenhui Liao, Zening Lin, Xinhong Chen, Chongyu Liu, Yuyi Zhang, and Lianwen Jin. 2023. Exploring ocr capabilities of gpt-4v (ision): A quantitative and in-depth evaluation. *arXiv preprint arXiv:2310.16809*.

Brandon Smock, Rohith Pesala, and Robin Abraham. 2022. Pubtables-1m: Towards comprehensive table extraction from unstructured documents. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4634–4642.

Granite Vision Team, Leonid Karlinsky, Assaf Arbelle, Abraham Daniels, Ahmed Nassar, Amit Alfassi, Bo Wu, Eli Schwartz, Dhiraj Joshi, Jovana Kondic, et al. 2025. Granite vision: a lightweight, open-source multimodal model for enterprise intelligence. *arXiv preprint arXiv:2502.09927*.

Chien-Yao Wang, Hong-Yuan Mark Liao, Yueh-Hua Wu, Ping-Yang Chen, Jun-Wei Hsieh, and I-Hau Yeh. 2020. Cspnet: A new backbone that can enhance learning capability of cnn. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 390–391.

Zhengyuan Yang, Linjie Li, Kevin Lin, Jianfeng Wang, Chung-Ching Lin, Zicheng Liu, and Lijuan Wang. 2023. The dawn of lmms: Preliminary explorations with gpt-4v (ision). *arXiv preprint arXiv:2309.17421*, 9(1):1.

Yuan Yao, Tianyu Yu, Ao Zhang, Chongyi Wang, Junbo Cui, Hongji Zhu, Tianchi Cai, Haoyu Li, Weilin Zhao, Zhihui He, et al. 2024. Minicpm-v:

A gpt-4v level mllm on your phone. *arXiv preprint arXiv:2408.01800*.

Jiaquan Ye, Xianbiao Qi, Yelin He, Yihao Chen, Dengyi Gu, Peng Gao, and Rong Xiao. 2021. Pingan-vcgroup's solution for icdar 2021 competition on scientific literature parsing task b: table recognition to html. *arXiv preprint arXiv:2105.01848*.

Guanghua Yu, Qinyao Chang, Wenyu Lv, Chang Xu, Cheng Cui, Wei Ji, Qingqing Dang, Kaipeng Deng, Guanzhong Wang, Yuning Du, et al. 2021. Pp-picodet: A better real-time object detector on mobile devices. *arXiv preprint arXiv:2111.00902*.

Xu Zhong, Elaheh ShafieiBavani, and Antonio Jimeno Yepes. 2020. Image-based table recognition: data, model, and evaluation. In *European conference on computer vision*, pages 564–580. Springer.

A Quantitative comparison with VLMs

Model	Finance		Marketing		PubTabNet		Sparse	
	Number of Samples							
	10	50	10	50	10	50	10	50
Llama Vision 3.2	53.80	43.17	37.02	41.22	49.83	46.31	23.23	30.66
Granite Vision 3.2	76.30	72.40	58.54	58.82	81.04	80.04	46.04	40.10
SLANet-1M (ours)	**99.50**	**99.48**	**99.78**	**99.16**	**99.92**	**99.55**	**97.69**	**99.20**

Table 8: Quantitative results (S-TEDS) comparison between Llama Vision, Granite Vision and SLANet-1M.

We selected two newly available large vision-language models (VLMs), Granite Vision 3.2 (Team et al., 2025) and Llama Vision 3.2 (AI, 2025), to compare quantitatively against SLANet-1M. We also evaluated MiniCPM-v (Yao et al., 2024), but its performance was insufficient for inclusion in the final comparison.

Following the methodology from (Peng et al., 2024), we randomly sampled a few images from each subset of the SynthTabNet dataset and conducted two experiments. In the first, we selected 10 images per subset; in the second, 50 images per subset. For each image, the VLMs were prompted with: *"Based on the table in the image, please generate the corresponding HTML code. Output only the HTML code."* We then computed the S-TEDS score for each output.

The results, shown in Table 8, clearly demonstrate that SLANet-1M significantly outperforms both Llama Vision and Granite Vision. Notably, while Granite Vision exhibited the strongest performance among the tested VLMs, it struggled considerably when processing large, information-dense tables.

B More Qualitative Results

Figures 7 and 8 present a qualitative comparison between SLANet-1M and Granite Vision. Since Granite Vision showed the best quantitative performance among the VLMs we evaluated, we chose it for a more in-depth qualitative analysis.

In Figure 7, panel (a) shows the input image, which comes from the PubTabNet subset of the SynthTabNet test set. Panel (b) displays the output of SLANet-1M, which achieves a perfect S-TEDS score of 1.00. While a few minor content errors are visible in some cells, these are attributable to limitations in the external models used for text detection and recognition, not SLANet-1M itself. Panel (c) shows Granite Vision's output, with a significantly lower S-TEDS score of 0.7658. The model incorrectly merges some cells, produces the wrong number of columns, and introduces an excess of blank cells.

In Figure 8, panel (a) shows an input image taken from the Finance subset of the SynthTabNet test set. Once again, SLANet-1M achieves a perfect S-TEDS score of 1.00, as shown in panel (b). In this case, Granite Vision in panel (c) performs noticeably better than in the previous example, though still not at SLANet-1M's level. This improvement can be attributed to the simpler and less structured layout of the input table.

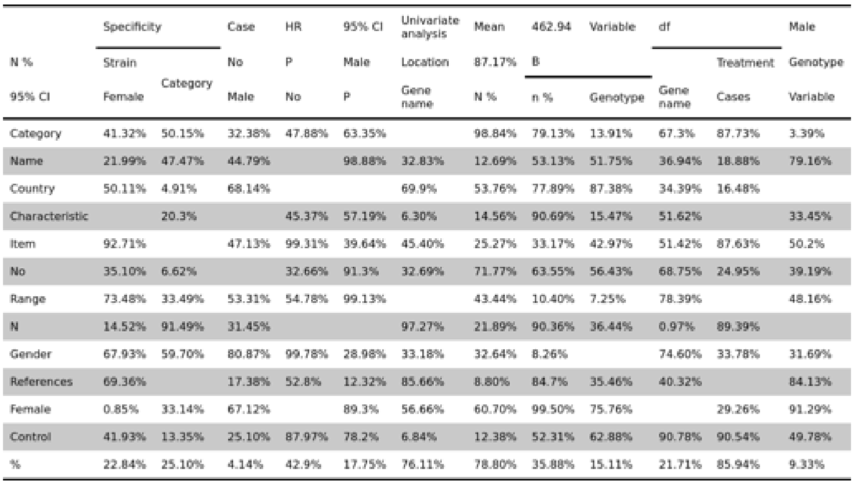
These qualitative results further support the superiority of SLANet-1M over some of the most recent VLMs in handling complex table understanding tasks.

C SLANet vs SLANet* vs SLANet-1M on PubTabNet

Models	**TEDS**	**S-TEDS**
SLANet (Li et al., 2022)	**95.89**	97.01
SLANet* (ours)	95.83	**97.48**
SLANet-1M (ours)	95.77	97.36

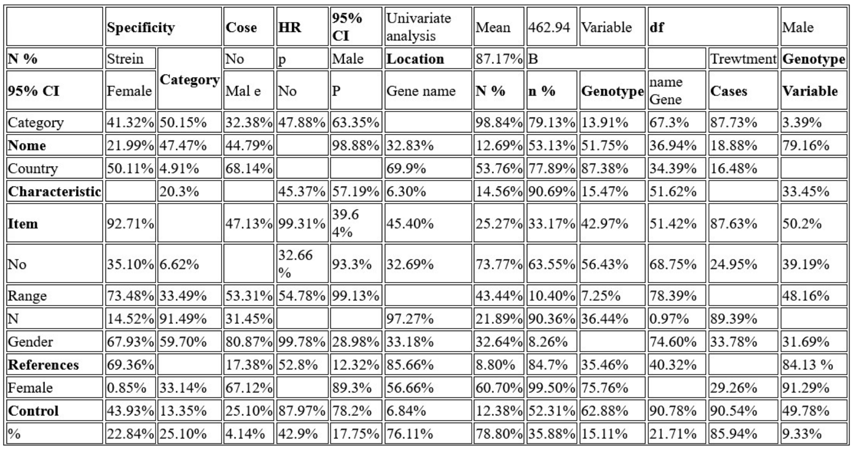
Table 9: Comparison on PubTabNet of models based on TEDS, S-TEDS.

Table 9 shows that SLANet-1M underperforms SLANet* on PubTabNet, likely due to SLANet* overfitting on the PubTabNet validation set, which was the only validation set used during its training.



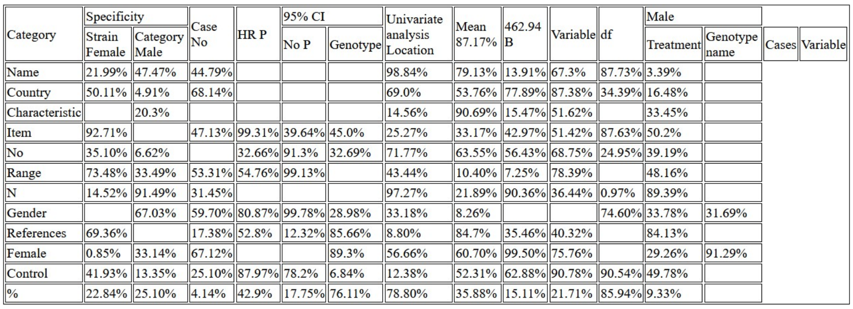
	Specificity		Case	HR	95% CI	Univariate analysis	Mean	462.94	Variable	df		Male
N %	Strain	Category	No	P	Male	Location	87.17%	B			Treatment	Genotype
95% CI	Female		Male	No	P	Gene name	N %	n %	Genotype	Gene name	Cases	Variable
Category	41.32%	50.15%	32.38%	47.88%	63.35%		98.84%	79.13%	13.91%	67.3%	87.73%	3.39%
Name	21.99%	47.47%	44.79%		98.88%	32.83%	12.69%	53.13%	51.75%	36.94%	18.88%	79.16%
Country	50.11%	4.91%	68.14%			69.9%	53.76%	77.89%	87.38%	34.39%	16.48%	
Characteristic		20.3%		45.37%	57.19%	6.30%	14.56%	90.69%	15.47%	51.62%		33.45%
Item	92.71%		47.13%	99.31%	39.64%	45.40%	25.27%	33.17%	42.97%	51.42%	87.63%	50.2%
No	35.10%	6.62%		32.66%	91.3%	32.69%	71.77%	63.55%	56.43%	68.75%	24.95%	39.19%
Range	73.48%	33.49%	53.31%	54.78%	99.13%		43.44%	10.40%	7.25%	78.39%		48.16%
N	14.52%	91.49%	31.45%			97.27%	21.89%	90.36%	36.44%	0.97%	89.39%	
Gender	67.93%	59.70%	80.87%	99.78%	28.98%	33.18%	32.64%	8.26%		74.60%	33.78%	31.69%
References	69.36%		17.38%	52.8%	12.32%	85.66%	8.80%	84.7%	35.46%	40.32%		84.13%
Female	0.85%	33.14%	67.12%		89.3%	56.66%	60.70%	99.50%	75.76%		29.26%	91.29%
Control	41.93%	13.35%	25.10%	87.97%	78.2%	6.84%	12.38%	52.31%	62.88%	90.78%	90.54%	49.78%
%	22.84%	25.10%	4.14%	42.9%	17.75%	76.11%	78.80%	35.88%	15.11%	21.71%	85.94%	9.33%

(a) Input table image (From PubTabNet subset of SynthTabNet).



	Specificity		**Cose**	**HR**	**95% CI**	Univariate analysis	Mean	462.94	Variable	**df**		Male
N %	Strein	**Category**	No	p	Male	**Location**	87.17%	B			Trewtment	**Genotype**
95% CI	Female		Mal e	No	P	Gene name	**N %**	**n %**	**Genotype**	name Gene	**Cases**	**Variable**
Category	41.32%	50.15%	32.38%	47.88%	63.35%		98.84%	79.13%	13.91%	67.3%	87.73%	3.39%
Nome	21.99%	47.47%	44.79%		98.88%	32.83%	12.69%	53.13%	51.75%	36.94%	18.88%	79.16%
Country	50.11%	4.91%	68.14%			69.9%	53.76%	77.89%	87.38%	34.39%	16.48%	
Characteristic		20.3%		45.37%	57.19%	6.30%	14.56%	90.69%	15.47%	51.62%		33.45%
Item	92.71%		47.13%	99.31%	39.6 4%	45.40%	25.27%	33.17%	42.97%	51.42%	87.63%	50.2%
No	35.10%	6.62%		32.66 %	93.3%	32.69%	73.77%	63.55%	56.43%	68.75%	24.95%	39.19%
Range	73.48%	33.49%	53.31%	54.78%	99.13%		43.44%	10.40%	7.25%	78.39%		48.16%
N	14.52%	91.49%	31.45%			97.27%	21.89%	90.36%	36.44%	0.97%	89.39%	
Gender	67.93%	59.70%	80.87%	99.78%	28.98%	33.18%	32.64%	8.26%		74.60%	33.78%	31.69%
References	69.36%		17.38%	52.8%	12.32%	85.66%	8.80%	84.7%	35.46%	40.32%		84.13 %
Female	0.85%	33.14%	67.12%		89.3%	56.66%	60.70%	99.50%	75.76%		29.26%	91.29%
Control	43.93%	13.35%	25.10%	87.97%	78.2%	6.84%	12.38%	52.31%	62.88%	90.78%	90.54%	49.78%
%	22.84%	25.10%	4.14%	42.9%	17.75%	76.11%	78.80%	35.88%	15.11%	21.71%	85.94%	9.33%

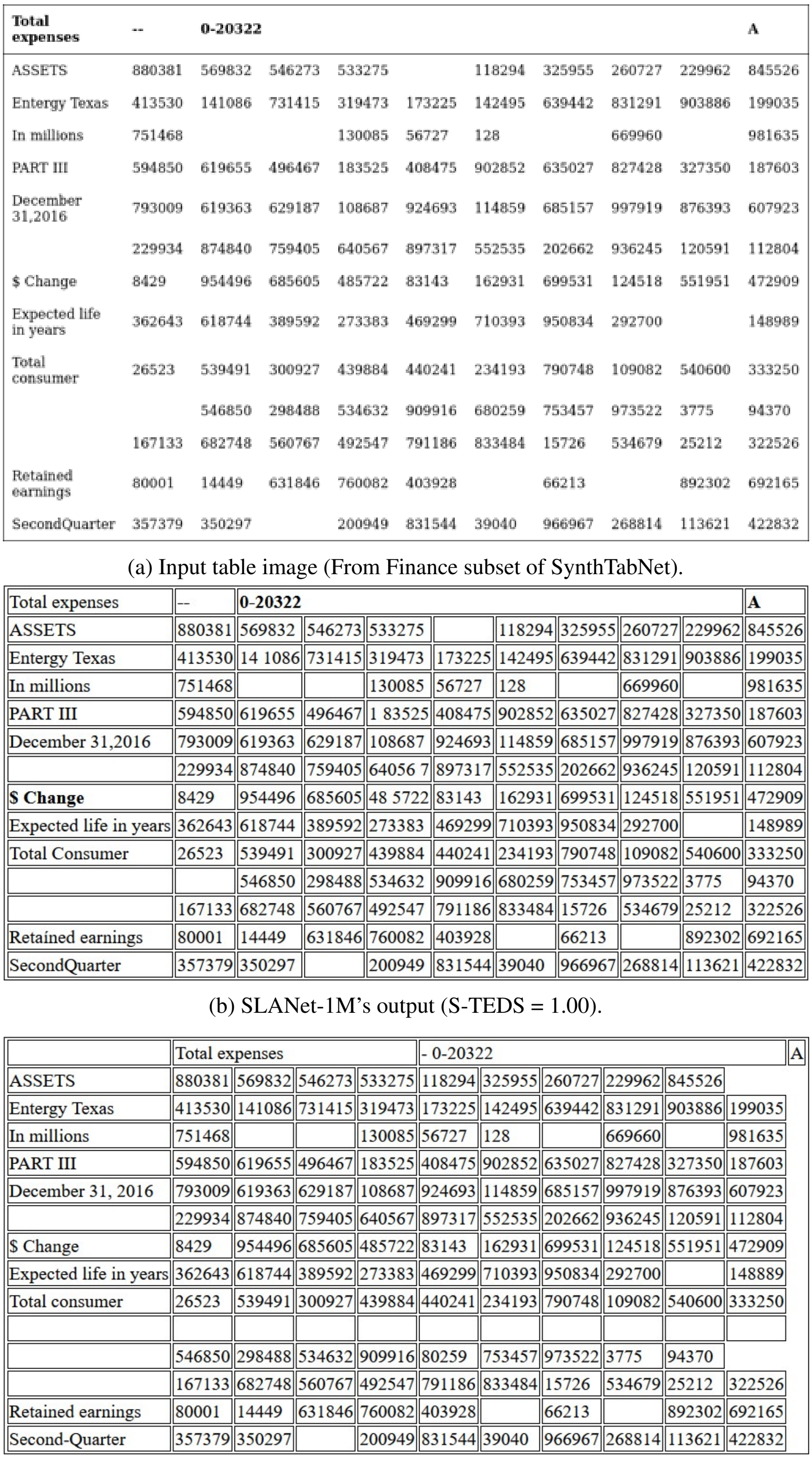
(b) SLANet-1M's output (S-TEDS = 1.00).



Category	Specificity		Case No	HR P	95% CI		Univariate analysis Location	Mean 87.17%	462.94 B	Variable	df	Male			
	Strain Female	Category Male			No P	Genotype						Treatment	Genotype name	Cases	Variable
Name	21.99%	47.47%	44.79%				98.84%	79.13%	13.91%	67.3%	87.73%	3.39%			
Country	50.11%	4.91%	68.14%				69.0%	53.76%	77.89%	87.38%	34.39%	16.48%			
Characteristic		20.3%					14.56%	90.69%	15.47%	51.62%		33.45%			
Item	92.71%		47.13%	99.31%	39.64%	45.0%	25.27%	33.17%	42.97%	51.42%	87.63%	50.2%			
No	35.10%	6.62%		32.66%	91.3%	32.69%	71.77%	63.55%	56.43%	68.75%	24.95%	39.19%			
Range	73.48%	33.49%	53.31%	54.76%	99.13%		43.44%	10.40%	7.25%	78.39%		48.16%			
N	14.52%	91.49%	31.45%				97.27%	21.89%	90.36%	36.44%	0.97%	89.39%			
Gender		67.03%	59.70%	80.87%	99.78%	28.98%	33.18%	8.26%			74.60%	33.78%	31.69%		
References	69.36%		17.38%	52.8%	12.32%	85.66%	8.80%	84.7%	35.46%	40.32%		84.13%			
Female	0.85%	33.14%	67.12%			89.3%	56.66%	60.70%	99.50%	75.76%		29.26%	91.29%		
Control	41.93%	13.35%	25.10%	87.97%	78.2%	6.84%	12.38%	52.31%	62.88%	90.78%	90.54%	49.78%			
%	22.84%	25.10%	4.14%	42.9%	17.75%	76.11%	78.80%	35.88%	15.11%	21.71%	85.94%	9.33%			

(c) Granite Vision's output (S-TEDS = 0.7658).

Figure 7: Qualitative comparison between Granite Vision and SLANet-1M.



Total expenses	**--**	**0-20322**								**A**
ASSETS	880381	569832	546273	533275		118294	325955	260727	229962	845526
Entergy Texas	413530	141086	731415	319473	173225	142495	639442	831291	903886	199035
In millions	751468			130085	56727	128		669960		981635
PART III	594850	619655	496467	183525	408475	902852	635027	827428	327350	187603
December 31,2016	793009	619363	629187	108687	924693	114859	685157	997919	876393	607923
	229934	874840	759405	640567	897317	552535	202662	936245	120591	112804
$ Change	8429	954496	685605	485722	83143	162931	699531	124518	551951	472909
Expected life in years	362643	618744	389592	273383	469299	710393	950834	292700		148989
Total consumer	26523	539491	300927	439884	440241	234193	790748	109082	540600	333250
		546850	298488	534632	909916	680259	753457	973522	3775	94370
	167133	682748	560767	492547	791186	833484	15726	534679	25212	322526
Retained earnings	80001	14449	631846	760082	403928		66213		892302	692165
SecondQuarter	357379	350297		200949	831544	39040	966967	268814	113621	422832

(a) Input table image (From Finance subset of SynthTabNet).

Total expenses	--	**0-20322**								**A**
ASSETS	880381	569832	546273	533275		118294	325955	260727	229962	845526
Entergy Texas	413530	14 1086	731415	319473	173225	142495	639442	831291	903886	199035
In millions	751468			130085	56727	128		669960		981635
PART III	594850	619655	496467	1 83525	408475	902852	635027	827428	327350	187603
December 31,2016	793009	619363	629187	108687	924693	114859	685157	997919	876393	607923
	229934	874840	759405	64056 7	897317	552535	202662	936245	120591	112804
$ Change	8429	954496	685605	48 5722	83143	162931	699531	124518	551951	472909
Expected life in years	362643	618744	389592	273383	469299	710393	950834	292700		148989
Total Consumer	26523	539491	300927	439884	440241	234193	790748	109082	540600	333250
		546850	298488	534632	909916	680259	753457	973522	3775	94370
	167133	682748	560767	492547	791186	833484	15726	534679	25212	322526
Retained earnings	80001	14449	631846	760082	403928		66213		892302	692165
SecondQuarter	357379	350297		200949	831544	39040	966967	268814	113621	422832

(b) SLANet-1M’s output (S-TEDS = 1.00).

	Total expenses				- 0-20322					A
ASSETS	880381	569832	546273	533275	118294	325955	260727	229962	845526	
Entergy Texas	413530	141086	731415	319473	173225	142495	639442	831291	903886	199035
In millions	751468			130085	56727	128		669660		981635
PART III	594850	619655	496467	183525	408475	902852	635027	827428	327350	187603
December 31, 2016	793009	619363	629187	108687	924693	114859	685157	997919	876393	607923
	229934	874840	759405	640567	897317	552535	202662	936245	120591	112804
$ Change	8429	954496	685605	485722	83143	162931	699531	124518	551951	472909
Expected life in years	362643	618744	389592	273383	469299	710393	950834	292700		148889
Total consumer	26523	539491	300927	439884	440241	234193	790748	109082	540600	333250
	546850	298488	534632	909916	80259	753457	973522	3775	94370	
	167133	682748	560767	492547	791186	833484	15726	534679	25212	322526
Retained earnings	80001	14449	631846	760082	403928		66213		892302	692165
Second-Quarter	357379	350297		200949	831544	39040	966967	268814	113621	422832

(c) Granite Vision’s output (S-TEDS = 0.9070).

Figure 8: Qualitative comparison between Granite Vision and SLANet-1M.