International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Integrative Deep Learning Strategy for Table Structure Classification and Recognition

Sridhar Patthi*[1], Sivaneasan Bala[2], T Sathish Kumar[3], Prasun Chakrabarti[4]

Submitted:13/03/2024 **Revised**: 28/04/2024 **Accepted**: 05/05/2024

Abstract: This paper discusses a unique method that takes advantage of deep neural networks to identify tables in documents. Traditional methods of table detection rely on dataset-specific heuristics that are prone to mistake. These existing methods are giving leverages to data for identifying tables with any arrangement. Table-structure recognition (TSR) is the arch focus of the investigation to identify the technique of digital document with table cell specification reproducibility and replicability. The proposed model works in identifying and specifying the extraction of table cell in structured, semi-structured and unstructured visual formats. Tables are crucial in presenting structural and semi-structural data which help in retrieval and analysis in databases. Table recognition involves identifying both the logical structure (cell relationships and spanning) and the physical structure (bounding boxes or cell content locations). Existing methods excel at appropriate predicting logical structures but at a high struggle with accurate physical structures like bounding boxes, which are vital for tasks like text extraction or table quality assurance. This proposal introduces a sequential coordinate decoding approach to enhance the accuracy of bounding box predictions by incorporating more visual information. While the coordinate sequence decoder provides a global context by leveraging the logical structure decoder's representation. This is deficient in local visual details, and is essential for accurate bounding box predictions made. This work with deep learning techniques solves the discussed challenge with an accuracy of 88.75%.

Keywords: *Classification, structures, endorsement, learning, machine intelligence, machine learning, machine vision, image identification*

1. Introduction

Table format information extraction from documents, table structure pinpointing is a critical task in the document analysis. Automating the procedure had made an enhanced significantly with the institution of deep learning. This research work examines the deep learning techniques to recognize table structures, emphasizing the assistance, drawbacks and possibilities [1]. The performance of these approaches, talk about how they have evolved, and designate ongoing problems in the field. Recapitulation on potential future study areas is included in the review's conclusion, with a focal point on the necessity of reliable and comprehensive models. Accurately recognizing the logical and physical structures of tables from unstructured photographs/documents/pdfs is a challenge in table recognition. Current techniques [2] suffer from imprecise bounding boxes for cells because there is insufficient local visual information in the coherent portrayal. Since accurate bounding boxes are essential to activities such as text extraction [3] and table analysis a major difficulty in table identification is observed and to be addressed. Current techniques accurately foresee the logical structure of tables, cell relationships and spanning. But it's still difficult to identify the table physical structure, such as bounding accurate boxes. Bounding box predictions are difficult as logical representation lacks local visual details. This emphasizes the necessity for more visual information to be incorporated into algorithms for better accuracy [4]. While most prior methods are restricted to Portable Data Frames, the suggested table detection technique operates directly on pictures, making it broadly applicable to any format. Deformable convolution can reshape its receptive field to match its input by adjusting it based just on the input. This alteration to the receptive field allows the network to support tables in any configuration.

The process of transforming semi-structured data [5] into a structured format that is machine-process able requires the use of table structure recognition (TSR). Heuristics and rule-based approaches have historically been used for this task, although they frequently have trouble with intricate table layouts. Deep learning's enhancement has revolutionized TSR identification to work precise and adaptable. This research presents a thorough analysis of the current state-of-the-art in deep learning-based TSR. Tables are included in different types of papers, including articles, financial reports, journals and scientific studies. Although they are very succinct and well-structured, their varied layouts are quite difficult for automated structure extraction to handle. For applications like knowledge extractions, data mining, and information retrieval, accurate TSR is essential. Rule-based methods are platformers to deep learning because they latter's the capacity to comprehend intricate patterns and apply generalizations to various table forms. TSR uses

[1]*Marri Laxman Reddy Institute of Technology & Management, Hyderabad, India*
ORCID ID: 0000-0002-1469-0204
[2] *Engineering Cluster, Singapore Institute of Technology, Singapore*
ORCID ID: 0000-0002-4271-2677
[3] *Hyderabad Institute of Technology and Management, Hyderabad, India*
ORCID ID: 0000-0002-8724-7166
[4] *Sir Padampat Singhania University, Rajastan, India*
ORCID ID: 0000-0001-8062-4144

deep learning models and methodologies.

Table detection in documents is an important task in document analysis and its understanding, with applications in information extraction, data mining, and automated document digitization. Deep learning models, including Convolutional Neural Networks (CNNs) [6], Recurrent Neural Networks (RNNs), and Vision Transformers (ViT), have shown promise in this domain. This document outlines the key deep learning techniques and equations used for detecting tables in various document formats.

1.1. Convolutional Neural Networks (CNNs):

CNNs' works more prowess in detecting and image processing tasks, which has led to their widespread use for TSR. They work especially well for identifying cells in a document, pdf, picture and table borders. The architecture of CNN is designed with innumerable convolutional layers and usually used in CNN-based methods. By pooling layers that extract features from the table image, tasks like region categorization, table boundary recognition, and cell detection are applied. The major limitations of CNNs are good at identifying table structures, but they frequently have trouble with intricate, nested tables.

$$y[i,j] = \sum_{m=1}^{M}\sum_{n=1}^{N} x[i+m, j+n].w[m,n] \quad (1)$$

Where term x[i,j] is the input, w[m,n] is the convolutional kernel or filter and y[i,j] is the output feature map Eqn. (1) . The CNN model for table detection typically uses multiple convolution layers, followed by pooling layers Eqn. (2) and a classification head to detect table regions in documents.

$$y[i,j] = max\ (x[i{:}\, i+p, j{:}\, j+p])] \quad (2)$$

Where ‘p’ is defined as pooling size.

1.2. Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM)

Long-Short Term Memory (LSTM) and Recurrent Neural Networks (RNNs) are employed in TSR to simulate the sequential nature of table structure as rows and columns. The architecture of LSTM &RNN are used for processing the table data [7] as sequences, these models are able to identify dependencies between different table cells, which can be used to apply in identifying the logical structure of tables, such as row and column headings, RNNs and LSTMs are especially helpful. Restrictions of LSTMs and RNNs may have problems with vanishing gradients and be computationally costly.

$$h_t = \sigma(W_{xh}x_t + W_{hh}h_{t-1} + b_h) \quad (3)$$

Where,

h_t is the hidden state at time step ‘t’

x_t is the input at time step ‘t’

W_{xh} and W_{hh} are weight matrices

b_h is the bias

σ is a non-linear activation function

RNNs are useful in sequence-based tasks, such as detecting tables in multipage documents where there is a temporal relationship between the sections. The hidden state of the RNN can capture contextual information about the table’s layout. The formulae for the layout is shown in Eqn. (3). Bidirectional RNNs (Bi-RNNs) are commonly used in document table detection to capture dependencies Eqn. (4a, 4b) from both past and future sections of the document.

$$\underset{h_t}{\rightarrow} = \sigma\ (W_{xh}x_t + W_{hh} \underset{h_{t-1}}{\longrightarrow} + b_h \quad (4a)$$

$$\underset{h_t}{\rightarrow} = \sigma\ (W_{xh}x_t + W_{hh} \underset{h_{t+1}}{\longleftarrow} + b_h \quad (4b)$$

1.3 Transformer Models

Transformers have demonstrated considerable covenant in TSR, especially for complex table structures, to their self-attention methods. The architecture of transformers works more accurate and able to capture long-range dependencies since they do not rely on sequential data processing [8]. The applications of transformers proven effective when used in multi-modal contexts (e.g., merging textual and visual input) for tasks like table parsing and structure recognition. The limitations of transformers are computationally demanding and require a lot of data for training.

1.4 Hybrid Models

Enhanced usage of each deep learning architecture's unique characteristics, recent research has concentrated on merging various architectures forming a single entity of model. CNN + RNN is used to manage intricate table layouts, this is in combination of CNNs' feature extraction powers and RNNs' sequential processing powers. Multi-Modal Approaches are used to increases recognition accuracy of various table structures by integrating textual and visual data [9] using CNNs and Transformers.

2. Dataset:

Table recognition performance of deep learning models is usually assessed using benchmark datasets like TableBank, PubTabNet, and ICDAR. Datasets offer a high test with the resilience of model because they encompass a wide range of table formats and document formats. Performance [10] are evaluated using metric og Precision, recall, F1-score, and Intersection over Union (IoU) for table detections. Metrics for structure recognition frequently centre on the precision of spanning cell identifications, row and column border detection. Deep learning models have been rapidly increasingly adopted for table recognition due to their superior abilities to handle complex patterns in both the structure and content of tables. These models can be broadly categorized into several types based on their architectures and the specific table recognitions tasks that they perform.

2.1 Table Detection

Initial stage of table recognitions is called table detection, and it involves locating table regions inside documents. Convolutional Neural Networks (CNNs), is one type of deep learning model that have proven to be quite successful in this task. This paper presents the enhancement of CNN with advancement of newer technologies [11].

- **CNN-Based Detection**: Rectangular sections in document and images can be identified by CNN’s, which are very apt at processing images and have been used to recognize tables. To enable these algorithm and work to discern between tables and other types of content, they are trained on extensive datasets of document pictures annotated with table region.
- **Object Detection Frameworks**: Table detections has made use of popular object detection frameworks such as YOLO (You Only Look Once) and Faster R-CNN. These models categorize the unidentified regions into several categories (e.g., simple tables, nested tables) in addition to detecting table boundary.

2.2 Table Structure Recognition

The process of determining a table's structure, mainly includes locating its rows, columns, and cell boundaries, after identifying the tables. This task is more difficult and necessitates are comprehending [12].

- **Grid-Based Approaches**: Grid-based techniques, is used in which the table is partitioned into a grid of cells, it employs CNNs. Each grid cell is subsequently assigned by the model to one of four categories like table cell, row header, column header, or empty space.
- **Sequence Models:** Tables' sequential characteristics has been modelled using recurrent neural networks (RNNs), especially Long Short-Term Memory (LSTM) networks. These models come extremely handy for tasks like determining rows and columns sequences, particularly in tables that have intricately spanned cells.
- **Transformer-Based Models**: Table structure recognitions has been implemented using transformers, which are well-known for their effectiveness in natural language processing. Complex tables with many layers of headers and spanning cells are especially well-suited to them due to their capacity to represent long-range dependencies.

2.3 Table Content Recognition

Extracting textual or numerical data from each cell and comprehending its significance within the table's context are necessary steps in identifying the content of the table. Recognizing table structures is a basic task in document analysis that allows structured data to be extracted from different types of documents. Table structure recognition algorithms function much better now that Transformers, with their strong self-attention mechanisms, are on the scene. This study discusses the design, uses, and benefits of Transformers over conventional techniques, including a thorough investigation of their application in table structure recognition. Additionally, the paper discusses the difficulties in using Transformer-based models and makes recommendations for possible future research avenues to increase their effectiveness in this field.

Optical Character Recognition (OCR): To extract text from table cells, deep learning models are frequently coupled with OCR engines. OCR is enhanced by deep learning, particularly when it comes to text recognition in noisy or low-resolutions in photos.

Contextual Understanding: Enhanced modelling comprehends the relationship between table cells by using contextual information. For instance, a tabular cell that is identified are identity as a header or data cell may be determined by its neighboring cell contents. In this situations, attention mechanisms and transformers are especially helpful.

2.4 Transformers:

Testimony of the rows, columns, and cells of tables contained in documents is known as table structure recognition (TSR), and it is essential step in transforming semi-structured data into a machine-readable format. When dealing with complex table layouts, traditional TSR method that rely on hand-crafted features or heuristic criteria have failed frequently. Transformers are capable to give resilient solutions to the problems presented by complex and varied table structures. Primitively created for natural language processing applications, transformers have been effectively adapted for TSR [13]. Transformers are a class of neural networks architecture where input data processing is done through self-attention processes. Transformers are better at capturing long-range dependencies since they do not handle data in a sequential manners like Recurrent Neural Networks (RNNs) do. Because of this, they are especially well-suited for tasks like table structure identifications, where comprehension of the relationships between the table's components are essential.

3. Review of research efforts:

Schreiber, Sebastian, et al. in this paper, a deep learning method for recognizing tables and organization in textual pictures is presented: DeepDeSRT. The method uses domain adaption and transfer learning methods to identify patterns and tables. The initial stage is to find table groupings in the written material, a process similar to finding objects in natural scenes. The method may also determine the columns as well as rows of a table. The authors scaled skip-pooled features and then incorporated FCN-2s method. To improve speed, they substituted scaled layers for standardized levels. Despite just one misinterpretation containing a non-table element, the approach produced good results with the ICDAR 2013 tabular issue set. Background pixel importance was brought up, and deep learning-focused linguistics partition research expertise was utilized to find patterns [17].

Siddiqui, Shoaib Ahmed, et al. presents a unique approach that makes use of flexible convolutional neural networks to analyze columnar patterns in textual pictures. For the transmission of important information, especially in economic and academic records, tabulated representations are essential. To tackle the issue of structure identification problem, the proposed DeepTabStR (Underground Tabular Structural Recognizer) employs a flexible convolutional method. Rather of using an established arrangement, the approach uses additional offsets that enable the underlying layer to adjust itself. The model is evaluated using the TabStructDB & ICDAR-13 structure of table identification [31] datasets after being previously trained using the dataset from ImageNet. The outcomes indicate that DeepTabStR was able to divide up rows as well as columns in a variety of texts with effectiveness [18].

Gilani, Azka, et al. proposed a quicker R-CNN is an extensive network that is used within a suggested approach enabling table identification in documents to determine the exact length among the text as well as white areas on the picture. In addition to producing size and longitudinal consistent zone recommendations [32], the method has one interface for identifying objects. Using a UNLV data set that consists of ten thousand photos that include different table configurations the research assessed the system. The technique improved accurate findings from 44% to 60.5%, outperforming commercialized algorithms in identifying the existence of table within intricate designs [19].

Hashmi, Khurram Azeem, et al. said Modern tables analyzing algorithms have advanced significantly in the past few years thanks to the use of deep-learning approaches to table recognition and architectural classification. The past ten years have seen a huge increase in interest in table comprehension, as well as the development in Deep-Neural-Networks multiple data-sets supporting tables identification, classification, & identification [20] have been made available. While conventional techniques are being employed to identify table using camera captured page pictures, methods involving deep learning could be applied to advance current tables evaluation methods. By utilizing the combination of several deep-learning principles using just released data-sets, the outcomes for tables classification & identification techniques could be enhanced further. Although it hasn't been studied in the field of table evaluation, learning through reinforcement presents a fascinating and potential avenue for tables recognition and classification for the years to come [21].

Paliwal, Shubham Singh, et al. presented TableNet represents a deep neural network which identifies structures and detects tables using already trained VGG-19 characteristics. It employs a multitasking learning strategy and rules-driven row collection in collecting information for table fields. Using this ICDAR-2013 information set, the algorithm beats previous sophisticated models and cutting-edge techniques. The method of transfer learning is made possible by its low adjustments required for generalization to different samples. As a result of its capacity to move knowledge between earlier objectives to more recent ones, TableNet is the initial framework that can resolve multiple objectives concurrently [22].

Nishida, Kyosuke, et al. described a novel structure for deep-neural networks called TabNet is intended for the categorization of table types. It is composed by a convolutional-neural networks which extracts lexical characteristics as well as an RNN (recurrent neural network) which records word combinations per every cell, producing 3D tables dimension information. The design captures excellent linguistic characteristics and shows how tables are structured. Based upon the scaled macro averaged F1 value underlying 91.05% of the 3,567 unidentified tables of information, the outcomes demonstrate that TabNet has the ability to record the layout of data [23].

Rashid, Sheikh Faisal, et al. proposed interpreting table is essential for deriving insights of digitized along with scanned texts; nevertheless, because of changeable designs, table detection is difficult. The model of neural networks is used to categorize phrases either into tables or non-table classifications based on their location and relationship to nearby components, which is one of the proposed methods for capturing table properties. Around 89% recognition of tables efficiency is attained following additional processing, which makes it appropriate for a variety of publications with various designs. Defining the boundaries of columns provides the sole restriction [24].

6. Evaluation and Insights gained

This proposal introduces a sequential coordinate decoding approach to enhance the accuracy of bounding box predictions by incorporating more visual information. The coordinate sequence decoder uses the representation from the logical structure decoder to provide a global context but lacks local visual details, which are crucial for precise bounding box predictions using deep-learning technique. The sequence of recognition of table in three different forms is developed with the classification of dataset by morning a neural net of sequence learning model. The sequence learning model works with the implementation of encoders to the input to detect the table by attention as shown in Eqn. (6).

$$Attention\ (Q, K, V) = softmax\ (\frac{QK^T}{\sqrt{d_k}})\mathrm{V}$$

(6)

Where:

- QK^Tdk\frac{Q K^T}{\sqrt{d_k}}dkQKT is the scaled dot-product of the query and key vectors (scaled by dk\sqrt{d_k}dk to prevent extremely large gradients when dkd_kdk is large).
- softmax\text{softmax}softmax ensures the attention scores sum to 1, thus forming a distribution.
- The result is multiplied by the value matrix VVV, so that each value is weighted according to its relevance.

Future research on Transformers in TSR should focus on several key areas.

Data Augmentation: Developing effective data augmentation techniques could help mitigate the data scarcity issue, allowing Transformers to be trained more robustly on diverse table structures.

Model Efficiency: Research into more efficient Transformer architectures, such as lightweight Transformers or hybrid models, could reduce the computational burden while maintaining accuracy.

Cross-Document Generalization: Improving the ability of Transformer models to generalize across different document formats and types will be crucial for broader application.

```
Initialize model parameters;
for each epoch do
    for each batch (X, y) do
        Forward pass: y_pred = model(X);
        Compute loss: L = BCE(y_pred, y);
        Backward pass: compute gradients;
        Update parameters with Adam optimizer;
    end
    Update learning rate schedule;
end
```

The Region-based Convolutional Neural Networks (R-CNN) are used for object detection, and table detection. The architecture uses a combination of CNNs to generate region proposals, then classifies and refines those regions.

YOLO is a real-time object detection algorithm. It divides the input image into a grid and predicts bounding boxes and class probabilities for each grid cell. YOLO can be adapted for detecting tables in documents. Bounding Box Prediction: YOLO directly predicts the bounding box (bx,by,bw,bh)(b_x, b_y, b_w, b_h)(bx, by,bw,bh), along with confidence score CCC and class probabilities P(ci)P(c_i)P(ci):

Where:

- CCC is the confidence score of the bounding box.
- P(ctable)P(c_{\text{table}})P(ctable) is the class probability of the "table" class.

1. Attention:

$$Attention\ (Q, K, V) = softmax\ (\frac{QK^T}{\sqrt{d_k}})\mathrm{V}$$

2. Multi-Head Attention

MultiHead (Q,K,V) = Contact(head1, …., $head_h$) W_0

3. Feed-Forward Network:

FFN (x) = max (0, xW1 + b1) W2 + b2

4. Layer Normalization

LayerNorm (x + SubLayer(x))

5. Positional Encoding

PE(pos,2i)=sin(pos/10000(2i/d)), PE(pos,2i+1) = cos(pos/10000(2i/d)),

The method of extracting data from sources of information for more analysis or archiving is known as data-extraction. The technique of organizing and standardizing important study features using data from presentations as well as publications is known as extracting information. It's an essential first step in determining the degree of prejudice when conducting private research & compiling their results. Data is frequently extracted from a predetermined set of areas during interventional, screening, or predictive systematic evaluations.

The workflow for table detection and extraction from digital documents begins with the input of a digital document, followed by the deletion of non-table objects. Features are extracted using models like VGG or RESNET, and the table is wrapped for further processing. The system recognizes tables through feature mapping, employing Deep ConvNet or ROI projection for regional localization. An SSD (Single Shot MultiBox Detector) is used to detect table components, which are then classified. Background subtraction and foreground segmentation refine the detection, and a deep learning classifier is applied to enhance table recognition. The recognized tables are organized into a master structure, with bounding boxes drawn around the detected tables. Finally, data is extracted from the tables, and the model predicts the structure and content of the table.

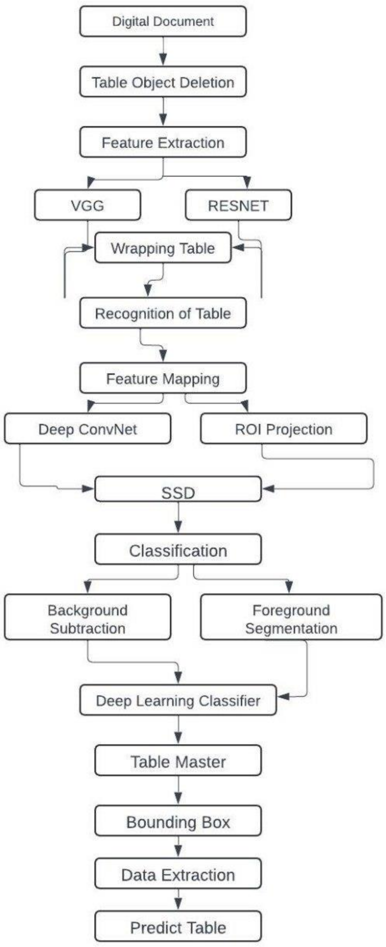


Fig1: Table-structure recognition (TSR) Workflow

Vision Transformers treat table cell as sequences of patches and apply transformer models, originally designed for NLP tasks, to process table structure. Each document tabular cell is divided into patches, which are embedded into a sequence of vectors. The patch embedding can be described in fig1 using the formula as in Eqn. (7).

$$E_{patch} Conv2D(X, kernel_{size} = p, stride = p) \quad (7)$$

Where X represents the input with the size(C,H,W), p is the patch size and Epatch defines the resulting patch embedding. The classification head outputs the probability of whether a table is present in the image. The classification token is added to the input, and the final output is as shown below in Eqn. (8).

$$y_{table} = \sigma(MLP(X_{CLS})) \quad (8)$$

Where XCLS embedded the process of classification token and sigmoid function.

Classification technique is used within statistical analysis along with recognition of table patterns, classification is a basic process which builds a tool for classification as well as provides a category labels to an occurrence given a collection specified features.

Classification represents a broad method for identifying, distinguishing, comprehending, along with organizing concepts and things onto classes, much like categorizing. It was extensively utilized in computing for tasks including making decisions, forecasting, and natural-language processing (NLP). Some of the most important activities in artificial intelligence involves the generation of classifier using sets of data containing pre-classified examples, typically referred to as training information. Training, also simply learning, is a method of modeling using data from training that is, creating a relationship from observable features of the table that accurately forecast the class of table cell from the given data. Background Subtraction is used within the fields of computer vision & processing of images, background subtraction of identified table features provides a method applied to identify objects that move in a series of table cells. To recognize table objects, it separates the background as well as foreground. Returning to the mask, it is unsuitable for static situations. Another method for identifying motion variations across successive frames is panel differentiation.

Foreground Segmentation uses classification techniques, foreground-segmentation is an automated vision approach which extracts foreground items out of an image's backdrop. For numerous researchers, the challenge of foreground-segmentation in moving films remains a difficult one. Numerous researchers researched a variety of conventionally created ways, nevertheless the use of those cutting-edge techniques hasn't produced positive outcomes. A substantial quantity of data with labels is used for training models based on deep learning. Information which can be categorized into distinct groups is known as labeled information. Deep-learning models go through a continuous training procedure. This implies that unless the algorithm merges on an outcome, it is refreshed with new information on a regular basis. Next, the algorithm predicts the name of the information point using its neural-network connections. Next, the predictive algorithm is adjusted to reduce the variance among the estimated as well as the real label after comparing its forecast and the label itself.

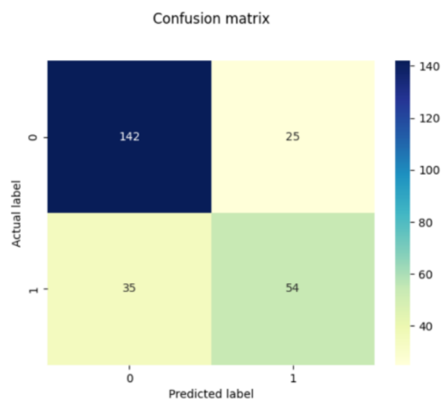
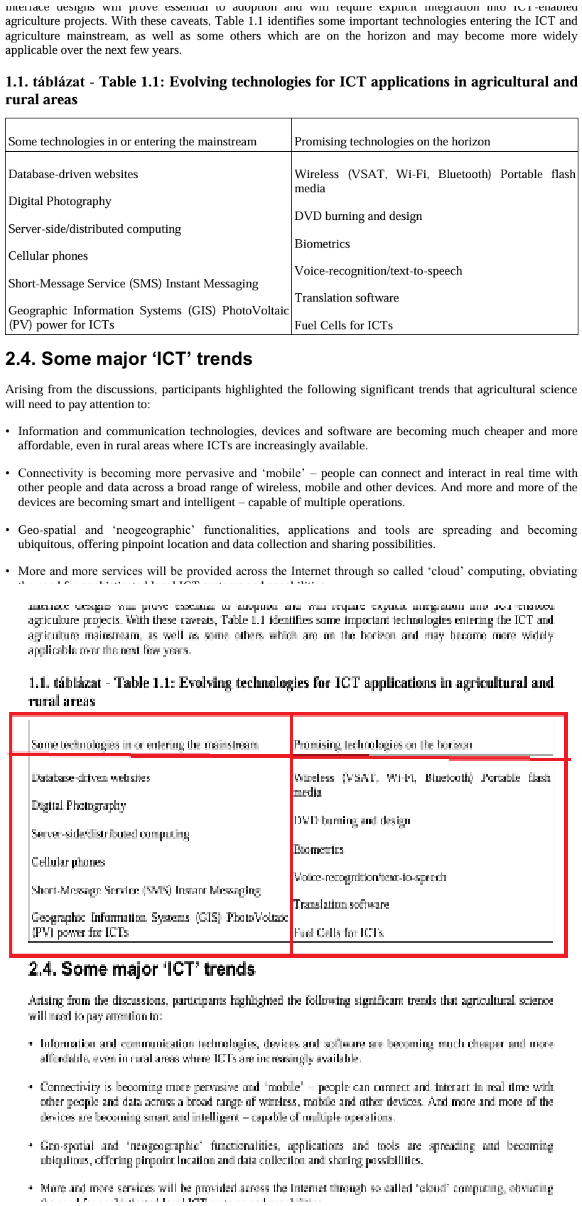


Fig2: TSR Confusion matrix

An item or collection of items in an electronic image are surrounded or encased by a rectangular shape (generally) is called a bounding box, frequently termed as a bounding-volume. A bounding-box is used to specify an item's position as well as dimensions in two or three dimensions, as well as to handling and analyzing the object. Bounding-boxes are used in many areas, including automation, processing images, & computer-vision. In TSR, bounding-boxes are utilized to recognize and classify the context and cell segmentation of the table. In processing the bounding-boxes are employed to row, column, and trimmed objects identification, this process is authenticated using the confusion matrix as shown in Fig2.


agriculture projects. With these caveats, Table 1.1 identifies some important technologies entering the ICT and agriculture mainstream, as well as some others which are on the horizon and may become more widely applicable over the next few years.

1.1. táblázat - Table 1.1: Evolving technologies for ICT applications in agricultural and rural areas

Some technologies in or entering the mainstream	Promising technologies on the horizon
Database-driven websites	Wireless (VSAT, Wi-Fi, Bluetooth) Portable flash media
Digital Photography	DVD burning and design
Server-side/distributed computing	Biometrics
Cellular phones	Voice-recognition/text-to-speech
Short-Message Service (SMS) Instant Messaging	Translation software
Geographic Information Systems (GIS) PhotoVoltaic (PV) power for ICTs	Fuel Cells for ICTs

2.4. Some major 'ICT' trends

Arising from the discussions, participants highlighted the following significant trends that agricultural science will need to pay attention to:

- Information and communication technologies, devices and software are becoming much cheaper and more affordable, even in rural areas where ICTs are increasingly available.
- Connectivity is becoming more pervasive and 'mobile' – people can connect and interact in real time with other people and data across a broad range of wireless, mobile and other devices. And more and more of the devices are becoming smart and intelligent – capable of multiple operations.
- Geo-spatial and 'neogeographic' functionalities, applications and tools are spreading and becoming ubiquitous, offering pinpoint location and data collection and sharing possibilities.
- More and more services will be provided across the Internet through so called 'cloud' computing, obviating

agriculture projects. With these caveats, Table 1.1 identifies some important technologies entering the ICT and agriculture mainstream, as well as some others which are on the horizon and may become more widely applicable over the next few years.

1.1. táblázat - Table 1.1: Evolving technologies for ICT applications in agricultural and rural areas

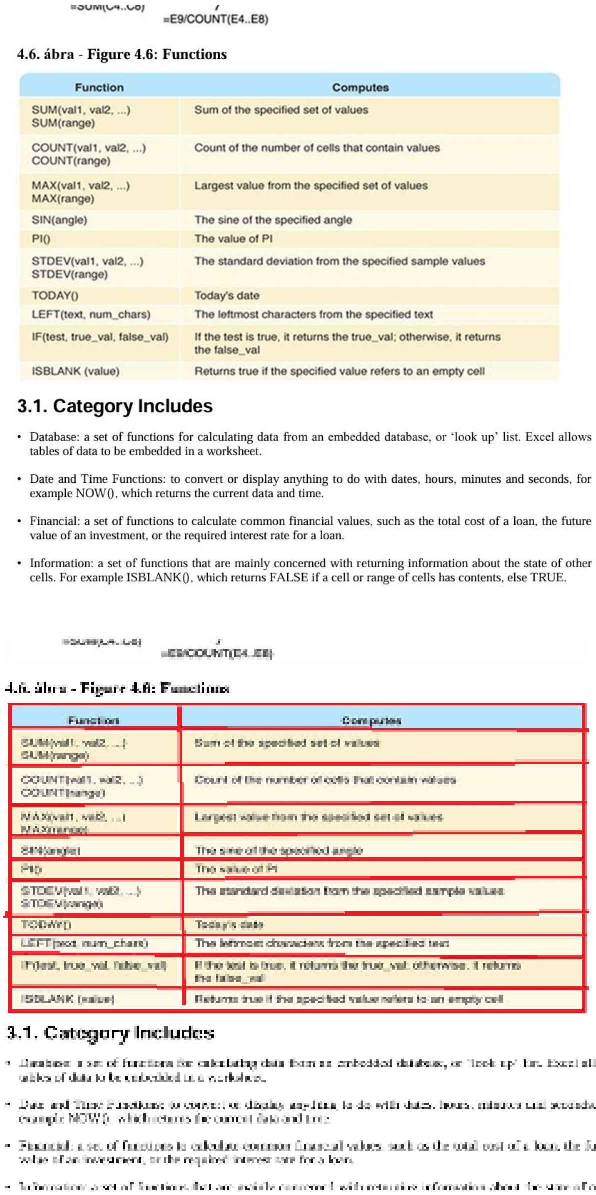
Some technologies in or entering the mainstream | Promising technologies on the horizon
Database-driven websites | Wireless (VSAT, Wi-Fi, Bluetooth) Portable flash media
Digital Photography | DVD burning and design
Server-side/distributed computing | Biometrics
Cellular phones | Voice-recognition/text-to-speech
Short-Message Service (SMS) Instant Messaging | Translation software
Geographic Information Systems (GIS) PhotoVoltaic (PV) power for ICTs | Fuel Cells for ICTs

2.4. Some major 'ICT' trends

Arising from the discussions, participants highlighted the following significant trends that agricultural science will need to pay attention to:

- Information and communication technologies, devices and software are becoming much cheaper and more affordable, even in rural areas where ICTs are increasingly available.
- Connectivity is becoming more pervasive and 'mobile' – people can connect and interact in real time with other people and data across a broad range of wireless, mobile and other devices. And more and more of the devices are becoming smart and intelligent – capable of multiple operations.
- Geo-spatial and 'neogeographic' functionalities, applications and tools are spreading and becoming ubiquitous, offering pinpoint location and data collection and sharing possibilities.
- More and more services will be provided across the Internet through so called 'cloud' computing, obviating

Fig3: Digital Document TSR Result


=E9/COUNT(E4..E8)

4.6. ábra - Figure 4.6: Functions

Function	Computes
SUM(val1, val2, ...) SUM(range)	Sum of the specified set of values
COUNT(val1, val2, ...) COUNT(range)	Count of the number of cells that contain values
MAX(val1, val2, ...) MAX(range)	Largest value from the specified set of values
SIN(angle)	The sine of the specified angle
PI()	The value of PI
STDEV(val1, val2, ...) STDEV(range)	The standard deviation from the specified sample values
TODAY()	Today's date
LEFT(text, num_chars)	The leftmost characters from the specified text
IF(test, true_val, false_val)	If the test is true, it returns the true_val; otherwise, it returns the false_val
ISBLANK (value)	Returns true if the specified value refers to an empty cell

3.1. Category Includes

- Database: a set of functions for calculating data from an embedded database, or 'look up' list. Excel allows tables of data to be embedded in a worksheet.
- Date and Time Functions: to convert or display anything to do with dates, hours, minutes and seconds, for example NOW(), which returns the current data and time.
- Financial: a set of functions to calculate common financial values, such as the total cost of a loan, the future value of an investment, or the required interest rate for a loan.
- Information: a set of functions that are mainly concerned with returning information about the state of other cells. For example ISBLANK(), which returns FALSE if a cell or range of cells has contents, else TRUE.

=E9/COUNT(E4..E8)

4.6. ábra - Figure 4.6: Functions

Function | Computes
SUM(val1, val2, ...) SUM(range) | Sum of the specified set of values
COUNT(val1, val2, ...) COUNT(range) | Count of the number of cells that contain values
MAX(val1, val2, ...) MAX(range) | Largest value from the specified set of values
SIN(angle) | The sine of the specified angle
PI() | The value of PI
STDEV(val1, val2, ...) STDEV(range) | The standard deviation from the specified sample values
TODAY() | Today's date
LEFT(text, num_chars) | The leftmost characters from the specified text
IF(test, true_val, false_val) | If the test is true, it returns the true_val; otherwise, it returns the false_val
ISBLANK (value) | Returns true if the specified value refers to an empty cell

3.1. Category Includes

Fig4: Image TSR Result

The workflow for table detection and extraction from digital documents begins with the input of a digital document, pdf and image as shown in fig3 and fig4 are followed by the deletion of non-table objects. Features are extracted using models like VGG or RESNET, and the table is wrapped for further processing. The system recognizes tables through feature mapping, employing Deep ConvNet or ROI projection for regional localization. An SSD (Single Shot MultiBox Detector) is used to detect table components, which are then classified. Background subtraction and foreground segmentation refine the detection, and a deep learning classifier is applied to enhance table recognition. The recognized tables are organized into a master structure, with bounding boxes drawn around the detected tables. Finally, data is extracted from the tables, and the model predicts the structure and content of the table. Table 1 below shows essential of accurate bounding box predictions made with a deep learning technique at an accuracy of 88.75%. Table 2 explains the different deep learning models which are recently applied for table detection and the proposed model which shows the metric wise analysis and improvement.

The Fig3a shows an organized table with minimum rows and columns without merged any merged columns or rows. The proposed model identifies the borders of the table in the document as shown in Fig3b. The Digital Document given as input is identified with the proper table structure recognition.

Fig 4a and 4b shows the Table Structure Recognition when the table consists of multiple rows and minimum columns. By these results, our proposed model has be recognising the tables from documents with maximum rows and maximum columns. The accuracy rate was high when compared to the existing deep learning models.

Table 1: Comparative Analysis *between performance metrics*

Model/ Metric	Precision	Recall	F1 Score	Bounding Box Accuracy
TableNet	84.2	82.5	83.3	81.2
Cascade TabNet	86.7	84.9	85.8	83.5
TabStructNet	85.9	83.6	84.7	82.8
Proposed Model	**88.75**	**87.1**	**87.9**	**88.75**

Table 2: Comparative Analysis: Deep Learning models for Table Detection between the existing and proposed

Metric/Aspect	TableNet	CascadeTabNet	TabStructNet	Proposed Model
Approach	Convolutional Neural Networks (CNNs) with U-Net architecture for table region and cell detection.	Cascade Mask R-CNN with additional attention mechanisms for enhanced detection.	Graph-based neural network leveraging logical and physical structure modeling.	Sequential coordinate decoding with global and local visual context integration.
Generalization Capability	Moderate; struggles with unstructured or complex table layouts.	High; excels in structured and semi-structured formats but struggles with unstructured formats.	High; designed for logical and physical structure extraction but limited local visual refinement.	**Very High; effective across structured, semi-structured, and unstructured table formats.**
Computational Efficiency	Efficient; requires less computational resources but sacrifices some accuracy.	Computationally intensive; suitable for high-performance systems.	Balanced; moderately computationally intensive with graph-based representations.	Computationally intensive; combines logical and visual contexts for high accuracy.
Use Cases	Suitable for detecting simple table layouts in structured documents.	Best for semi-structured tables; useful for applications like table QA and document analysis.	Focused on detailed logical relationships; excels in semi-structured but struggles with bounding box accuracy.	Robust for structured, semi-structured, and unstructured tables; ideal for text extraction, table QA, and database tasks.
Innovation	Introduced U-Net architecture for table detection in deep learning.	First to leverage attention mechanisms in a cascade R-CNN framework for tables.	Pioneered graph-based modeling for joint logical and physical structure recognition.	Hybrid decoding for enhanced bounding box predictions by integrating visual and logical contexts.

Conclusions

TSR introduces a sequential coordinate decoding approach to enhance the accuracy of bounding box predictions by incorporating more visual information with classification analysis of Table Structure through endorsement of deep learning. The coordinate sequence decoder provides a global context by leveraging the logical structure decoder's representation. The workflow for table detection and extraction from digital documents begins with the input of a digital document, pdf and image are followed by the deletion of non-table objects. Features are extracted using models like VGG or RESNET, and the table is wrapped for further processing. It is deficient in local visual details, which are essential for accurate bounding box predictions made with a deep learning technique at an accuracy of 88.75%.

References:

[1] Ting Chen, Saurabh Saxena, Lala Li, David J. Fleet, and Geoffrey Hinton. Pix2seq: A language modeling framework for object detection. In International Conference on Learning Representations, 2022. 4.

[2] Zewen Chi, Heyan Huang, HengDa Xu, Houjin Yu, Wanxuan Yin, and XianLing Mao. Complicated table structure recognition, 2019. 5, 6, 7

[3] St´ephane Clinchant, Herv´e D´ejean, Jean-Luc Meunier,Eva Maria Lang, and Florian Kleber. Comparing machine learning approaches for table recognition in historical register books. 2018 13th IAPR International Workshop on Document Analysis Systems (DAS), pages 133–138, 2018.

[4] Jifeng Dai, Yi Li, Kaiming He, and Jian Sun. R-fcn: Object detection via region-based fully convolutional networks. In Advances in Neural Information Processing Systems, pages 379–387. Curran Associates Inc., 2016.

[5] Yuntian Deng, Anssi Kanervisto, Jeffrey Ling, and Alexander M. Rush. Image-to-markup generation with coarse-to-fine attention. In Proceedings of the 34th International Conference on Machine Learning, pages 980–989, 2017.

[6] Yuntian Deng, David Rosenberg, and Gideon Mann. Challenges in end-to-end neural scientific table recognition. In 2019 International Conference on Document Analysis and Recognition, pages 894–901, 2019.

[7] Pascal Fischer, Alen Smajic, Giuseppe Abrami, and Alexander Mehler. Multi-type-td-tsr - extracting tables from document images using a multi-stage pipeline for table detection and table structure recognition: From ocr to structured table representations. In Stefan Edelkamp, Ralf M¨oller, and Elmar Rueckert, editors, KI 2021: Advances in Artificial Intelligence, pages 95–108. Springer International Publishing, 2021.

[8] Liangcai Gao, Yilun Huang, Herv´e D´ejean, Jean-Luc Meunier, Qinqin Yan, Yu Fang, Florian Kleber, and Eva Lang. Icdar 2019 competition on table detection and recognition (ct-dar). In 2019 International Conference on Document Analysis and Recognition, pages 1510–1515, 2019.

[9] Max G¨obel, Tamir Hassan, Ermelinda Oro, and Giorgio Orsi. A methodology for evaluating algorithms for table understanding in pdf documents. In Proceedings of the 2012 ACM Symposium on Document Engineering, page 45–48. Association for Computing Machinery, 2012.

[10] Zengyuan Guo, Yuechen Yu, Pengyuan Lv, Chengquan Zhang, Haojie Li, Zhihui Wang, Kun Yao, Jingtuo Liu, and Jingdong Wang. TRUST: An Accurate and End-to-End Table structure Recognizer Using Splitting-based Transformers, Aug. 2022.

[11] Max G¨obel, Tamir Hassan, Ermelinda Oro, and Giorgio Orsi. Icdar 2013 table competition. In 2013 12th International Conference on Document Analysis and Recognition, pages1449–1453, 2013.

[12] Kaiming He, Georgia Gkioxari, Piotr Dollar, and Ross Girshick. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Oct 2017.

[13] Antonio Jimeno-Yepes, Xu Zhong, and Douglas Burdick. Icdar 2021 competition on scientific literature parsing. arXive-prints, page arXiv:2106.14616, 2021.

[14] Saqib Ali Khan, Syed Khalid, Muhammad Ali Shahzad, and Faisal Shafait. Table structure extraction with bi-directional gated recurrent unit networks. In 2019 International Conference on Document Analysis and Recognition (ICDAR), pages 1366–1371, 2019.

[15] Enuji Lee, Jaewoo Park, Hyung Il Koo, and Nam Ik Cho. Deep-learning and graph-based approach to table structure recognition. Multimedia Tools and Applications, 81:5827-5848, 2022.

[16] Minghao Li, Lei Cui, Shaohan Huang, Furu Wei, Ming Zhou, and Zhoujun Li. TableBank: Table benchmark for image-based table detection and recognition. In Proceedings of the 12th Language Resources and Evaluation Conference, pages 1918–1925. European Language Resources Association, May 2020.

[17] Schreiber, Sebastian, et al. "Deepdesrt: Deep learning for detection and structure recognition of tables in document images." 2017 14th IAPR international conference on document analysis and recognition (ICDAR). Vol. 1. IEEE, 2017.

[18] Siddiqui, Shoaib Ahmed, et al. "Deeptabstr: Deep learning-based table structure recognition." 2019 international conference on document analysis and recognition (ICDAR). IEEE, 2019.

[19] Gilani, Azka, et al. "Table detection using deep learning." 2017 14th IAPR international conference on document analysis and recognition (ICDAR). Vol. 1. IEEE, 2017.

[20] Hashmi, Khurram Azeem, et al. "Current status and performance analysis of table recognition in document images with deep neural networks." IEEE Access 9

(2021): 87663-87685.

[21] Paliwal, Shubham Singh, et al. "Tablenet: Deep learning model for end-to-end table detection and tabular data extraction from scanned document images." 2019 International Conference on Document Analysis and Recognition (ICDAR). IEEE, 2019.

[22] Nishida, Kyosuke, et al. "Understanding the semantic structures of tables with a hybrid deep neural network architecture." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 31. No. 1. 2017.

[23] Rashid, Sheikh Faisal, et al. "Table recognition in heterogeneous documents using machine learning." 2017 14th IAPR International conference on document analysis and recognition (ICDAR). Vol. 1. IEEE, 2017.

[24] Hashmi, Khurram Azeem, et al. "Guided table structure recognition through anchor optimization." IEEE Access 9 (2021): 113521-113534.

[25] Shigarov, Alexey, Andrey Mikhailov, and Andrey Altaev. "Configurable table structure recognition in untagged PDF documents." Proceedings of the 2016 ACM symposium on document engineering. 2016.

[26] Siddiqui, Shoaib Ahmed, et al. "Rethinking semantic segmentation for table structure recognition in documents." 2019 international conference on document analysis and recognition (ICDAR). IEEE, 2019.

[27] Lin, Weihong, et al. "Tsrformer: Table structure recognition with transformers." Proceedings of the 30th ACM International Conference on Multimedia. 2022.