Engineering Applications of Artificial Intelligence 90 (2020) 103551

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Holistic design for deep learning-based discovery of tabular structures in datasheet images☆,☆☆

Ertugrul Kara [a], Mark Traquair [a], Murat Simsek [a,b], Burak Kantarci [a,*], Shahzad Khan [c]

[a] School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, ON, K1N 6N5, Canada
[b] Faculty of Aeronautics and Astronautics, Istanbul Technical University, Maslak, Istanbul, 34469, Turkey
[c] Lytica Inc., 308 Legget Dr, Kanata, ON K2K 1Y6, Canada

ARTICLE INFO

Keywords:
Deep learning
Image processing
Document processing
Table detection
Tabular data extraction
Page object detection
Structure detection

ABSTRACT

Extracting data from tabular structures contained within product datasheets is crucial in many contexts, particularly in the management and optimization of supply chains that serve various industries. In order to minimize human intervention, table detection and table structure detection form the essential functionality. However, a self-contained holistic solution to extract the tables as well as their columns and rows in not readily available. To address this challenge, This study presents a new formal procedure that consists of the following sequence: table detection, structure segmentation and holistic tabular structure detection on documents. The proposed table detection model outperforms the state-of-the-art solutions by achieving a recall value of 1.0 and a precision of more than 0.99 on public competition datasets. Furthermore, this work introduces a judging mechanism and an agreement-based post-processing procedure to incorporate hand-crafted rules into the deep learning models. Though the individual components achieve a new state-of-the-art F1-Score, when integrated the best achieved F-measure for the holistic system is 0.89.

1. Introduction

Computation power is becoming more accessible and affordable every day. With the advent of next generation wireless and mobile networking solutions as well as Artificial Intelligence (AI) and industry 4.0, Internet of Things (IoT) devices are becoming more pervasive. A global and integrated supply chain has arisen with innovative intelligent and connected manufacturing processes to provide the components and modules required to support rapid new product introduction in these industries (Vaidya et al., 2018). These new supply chains are information intensive and require digitization of traditional methods and the application of novel IoT, smart manufacturing, and cloud supported operations (Tjahjono et al., 2017). Supply chains are poised to benefit from the advances in AI. Governments, hospitals and other organizations that rely on non-digital data are positioned to benefit from digital disruption, enhanced information flows and intelligent supply chain methodologies. Product specification datasheet for supply chain is traditionally contained in either printed, written documents or online as PDF documents. The aforementioned organizations usually deal with millions of these documents. However, an important hurdle to complete digitization is that these documents are formatted for human consumption, and do not cater to the digitization requirements that are essential for efficient software mediated processing.

Modern supply chains consist of a number of manufacturers, suppliers, distributors, contract manufacturers and retailers. The importance in automation of document extraction, such as the process seen in Fig. 1, is continuously increasing. With the exponential growth in data volumes, extraction with human efforts becomes infeasible (Traquair et al., 2019; Gilani et al., 2017). Public datasets and competitions, such as ICDAR (Karatzas et al., 2013) and UNLV (Rice et al., 2012) have been created to systematically advance this popular and important task.

In this work, the focus is on the information discovery problems in documents required for supply chains. Suppliers, especially electronic component manufacturers, release their datasheets in a

☆ This work was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) ENGAGE Program and Ontario Centres of Excellence, Canada Voucher for Innovation and Productivity (VIP) I Program under Project Number 29931, Gnowit Inc, Lytica Inc, and Research Fund of the Istanbul Technical University, Turkey. Project Number: MUA-2019-41997.

☆☆ One or more of the authors of this paper have disclosed potential or pertinent conflicts of interest, which may include receipt of payment, either direct or indirect, institutional support, or association with an entity in the biomedical field which may be perceived to have potential conflict of interest with this work. For full disclosure statements refer to https://doi.org/10.1016/j.engappai.2020.103551.

* Corresponding author.
E-mail addresses: ekara044@uottawa.ca (E. Kara), mtraq059@uottawa.ca (M. Traquair), murat.simsek@uottawa.ca (M. Simsek), burak.kantarci@uottawa.ca (B. Kantarci), shahzad_khan@lytica.com (S. Khan).

https://doi.org/10.1016/j.engappai.2020.103551
Received 5 July 2019; Received in revised form 17 January 2020; Accepted 7 February 2020
Available online 15 February 2020
0952-1976/

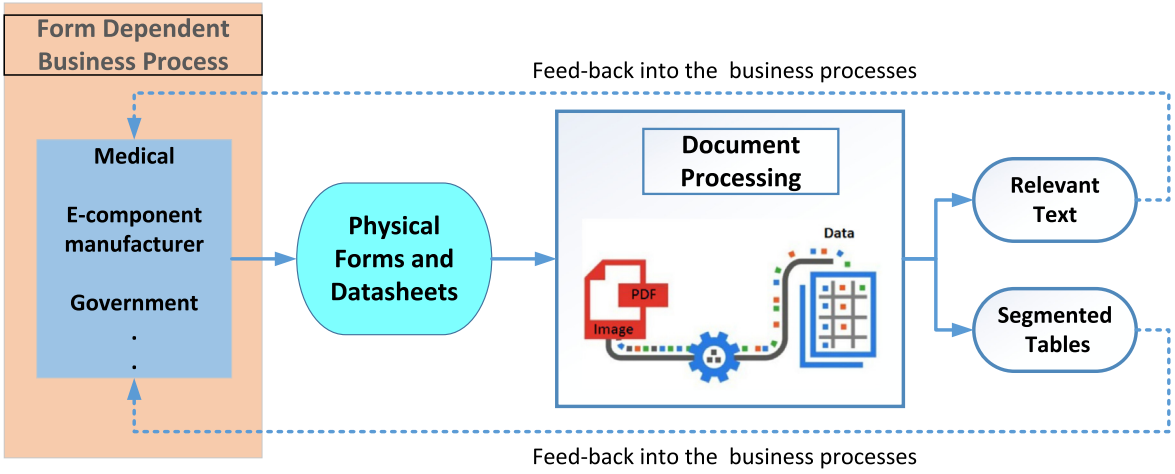


Fig. 1. Example document flow which exists throughout all industry.

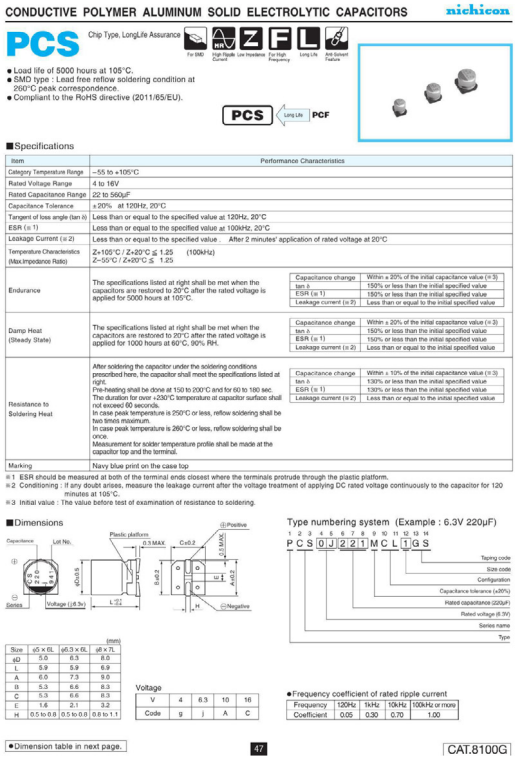
human-readable format. Crucial information such as sizes of the components, voltage values, etc. are presented in tabular structures. Some examples of the dataset can be seen in Fig. 2. These tabular structures include formulas, figures, and tables, that differ in formatting, structure, and content based on the publishing manufacturer, commodity type and product family. This is significant variety in the format that the data is presented in with the datasheet publisher's goal prioritizing aesthetics, information density, and cohesiveness of tables — with multiple tables employed to represent different facets of the product to make it comprehensible to a human reader in bite-sized chunks. Additionally, the quality of representation can vary based on the primary language of the publisher and their document preparation process sophistication. These factors are a solid argument for analyzing the documents as images rather than as MS Word documents or structured PDFs. Many commercially available tools require structured templates and text to be associated with the document, which is difficult to impose on the millions of manufacturers that constitute the global supply chain. This is the justification to rely solely on plain images as the system input.

Documents, particularly data sheets and the tables contained within them do not have consistent formatting with an absolute or authoritative set of rules to follow. There are innumerable variations to how vital data may be represented, with several papers focusing on their segmentation (Mao et al., 2003) or purely the detection of the tables contained within them (Fang et al., 2011; Schreiber et al., 2017; Tran et al., 2015; Traquair et al., 2019).

As the majority of the data sheets are often scanned documents, deep learning can offer viable solutions for handling the datasheet images. When attempting to extract meaning from tabular data embedded in documents, the following two challenges are faced: (1) table detection, and (2) segmentation of detected tables into semantically related portions (i.e. columns and rows). These two tasks introduce individual challenges, where table detection requires identifying the salient elements of the document, and extraction of the table object from the surrounding text. A recent study (Kara et al., 2019) involved training the Mask Region-Based Convolutional Neural Network (Mask R-CNN) as per (He et al., 2017) to divide the tables into relevant columns and rows. This task is significantly more challenging than the table detection due to the abstract nature of tables; indeed, the contents of a table itself involve cells that are merged, with split columns or rows, with figures contained therein, or even examples containing sub-tables within cells themselves. Performance of this system – when tested with perfectly formatted tables – can reach an Average Precision (AP) of 0.95 as presented in Kara et al. (2019), however, it is worth noting that the presented table structure detection system assumes perfect tables and is sensitive to errors in these tables. As the number of the documents processed are quite large, the number of tables detected by this system can overwhelm the ability of humans efforts to manually adjust and improve.

In this paper, the information extraction from tabular structures are broken down into two steps: (1) Detection of tables, (2) Detection of rows and columns of the detected tables. In both steps, deep learning is exploited, specifically Convolutional Neural Networks (CNNs) as it has been shown to perform well on detection tasks (Shin et al., 2016; He et al., 2017). To align with the literature, Mask-RCNN (He et al., 2017) is identified as the primary network as it was previously shown to outperform many other state-of-the-art deep learning methods in a similar but narrower problem (Kara et al., 2019). Internal table structure detection is a challenging task in the context of CNNs because objects (specifically rows) are usually only a few pixels in height. Considering long strides, large filters and pooling layers in CNN networks, this task introduces further complexity to the traditional CNN-based solutions (Hu and Ramanan, 2017). The previously available techniques focus on table detection and structure detection applied to full documents. The research study in this field usually focuses on only one of these problems in lieu of beneficially combining them. While approaching these problems, researchers either use heuristics-based methods or deep learning (Kieninger and Dengel, 1998; Oro and Ruffolo, 2009; Traquair et al., 2019; Kavasidis et al., 2018). Robust solutions for the table detection problem exists with object detection networks whereas the structure detection problem is usually addressed by being considered as a segmentation problem with Fully Convolutional Networks (Shelhamer et al., 2017). The contributions of this paper can be summarized as follows;

- This work proposes a novel holistic solution to extract tables, rows, and columns, which is also referred to as an End-to-End (E2E) detection system. The proposed holistic system is based on Mask-RCNN (He et al., 2017) which is an instance segmentation method equipped with Feature Pyramid Networks (Lin et al., 2017a) and Fully Convolutional Networks (Shelhamer et al., 2017). This paper provides useful insights on the usefulness of these architectural decisions particularly for the tabular structure detection problem. The proposed solution is to adopt a pre-trained Mask-RCNN model by He et al. (2017) and apply transfer learning for detection of tabular structures.
- This paper presents the impact of choosing to detect 'tabular regions' instead of 'tables'. When the training set is adapted to include all the tabular regions, detection performance increased immensely. Even though this may lead to false-positives when



CONDUCTIVE POLYMER ALUMINUM SOLID ELECTROLYTIC CAPACITORS

nichicon

PCS

Chip Type, LongLife Assurance

- Lead life of 5000 hours at 105°C.
- SMD type : Lead free reflow soldering condition at 260°C peak correspondence.
- Compliant to the RoHS directive (2011/65/EU).

PCS Long Life PCF

■Specifications

Item	Performance Characteristics
Category Temperature Range	−55 to +105°C
Rated Voltage Range	4 to 16V
Rated Capacitance Range	22 to 560µF
Capacitance Tolerance	±20% at 120Hz, 20°C
Tangent of loss angle (tan δ)	Less than or equal to the specified value at 120Hz, 20°C
ESR (≅ 1)	Less than or equal to the specified value at 100kHz, 20°C
Leakage Current (≅ 2)	Less than or equal to the specified value . After 2 minutes' application of rated voltage at 20°C
Temperature Characteristics (Max.Impedance Ratio)	Z+105°C / Z+20°C ≦ 1.25 (100kHz) Z−55°C / Z+20°C ≦ 1.25
Endurance	The specifications listed at right shall be met when the capacitors are restored to 20°C after the rated voltage is applied for 5000 hours at 105°C.
Damp Heat (Steady State)	The specifications listed at right shall be met when the capacitors are restored to 20°C after the rated voltage is applied for 1000 hours at 60°C, 90% RH.
Resistance to Soldering Heat	After soldering the capacitor under the soldering conditions prescribed here, the capacitor shall meet the specifications listed at right. Pre-heating shall be done at 150 to 200°C and for 60 to 180 sec. The duration for over +230°C temperature at capacitor surface shall not exceed 60 seconds. In case peak temperature is 250°C or less, reflow soldering shall be two times maximum. In case peak temperature is 260°C or less, reflow soldering shall be once. Measurement for solder temperature profile shall be made at the capacitor top and the terminal.
Marking	Navy blue print on the case top

Capacitance change	Within ± 20% of the initial capacitance value (≅ 3)
tan δ	150% or less than the initial specified value
ESR (≅ 1)	150% or less than the initial specified value
Leakage current (≅ 2)	Less than or equal to the initial specified value

Capacitance change	Within ± 20% of the initial capacitance value (≅ 3)
tan δ	150% or less than the initial specified value
ESR (≅ 1)	150% or less than the initial specified value
Leakage current (≅ 2)	Less than or equal to the initial specified value

Capacitance change	Within ± 10% of the initial capacitance value (≅ 3)
tan δ	130% or less than the initial specified value
ESR (≅ 1)	130% or less than the initial specified value
Leakage current (≅ 2)	Less than or equal to the initial specified value

≅1 ESR should be measured at both of the terminal ends closest where the terminals protrude through the plastic platform.
≅2 Conditioning : If any doubt arises, measure the leakage current after the voltage treatment of applying DC rated voltage continuously to the capacitor for 120 minutes at 105°C.
≅3 Initial value : The value before test of examination of resistance to soldering.

■Dimensions

Plastic platform

Type numbering system (Example : 6.3V 220µF)

P C S 0 J 2 2 1 M C L 1 G S

Taping code
Size code
Configuration
Capacitance tolerance (±20%)
Rated capacitance (220µF)
Rated voltage (6.3V)
Series name
Type

Size	φD × L	φ6.3 × 6L	φ8 × 7L
φD	5.0	6.3	8.0
L	5.9	5.9	6.9
A	6.0	7.3	9.0
B	5.3	6.6	8.3
C	5.3	6.6	8.3
E	1.6	2.1	3.2
H	0.5 to 0.8	0.5 to 0.8	0.6 to 1.1

(mm)

Voltage

V	4	6.3	10	16
Code	g	j	A	C

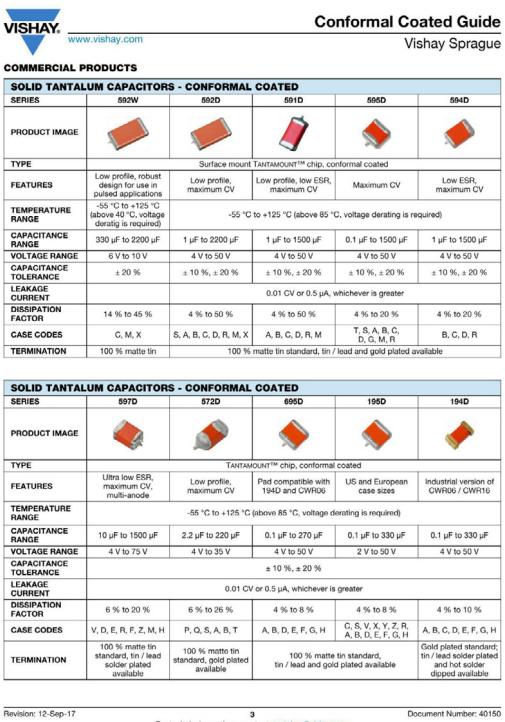
●Frequency coefficient of rated ripple current

Frequency	120Hz	1kHz	10kHz	100kHz or more
Coefficient	0.05	0.30	0.70	1.00

●Dimension table in next page.

47

CAT.8100G



VISHAY. www.vishay.com

Conformal Coated Guide

Vishay Sprague

COMMERCIAL PRODUCTS

SOLID TANTALUM CAPACITORS - CONFORMAL COATED

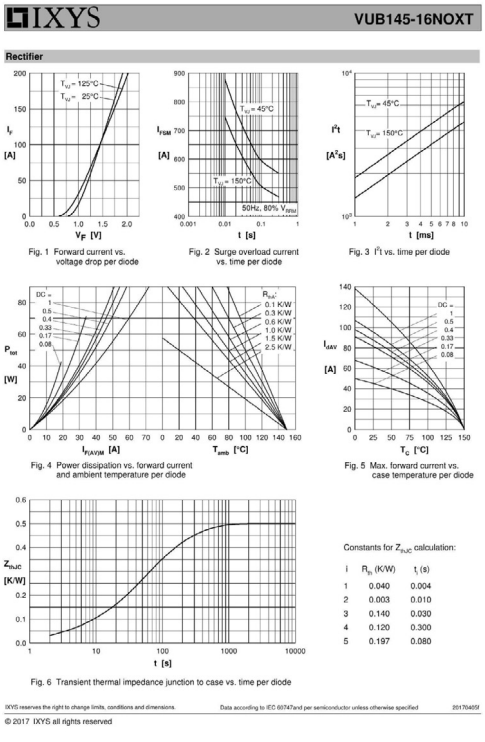
SERIES	592W	592D	591D	595D	594D
PRODUCT IMAGE					
TYPE	Surface mount TANTAMOUNT™ chip, conformal coated				
FEATURES	Low profile, robust design for use in pulsed applications	Low profile, maximum CV	Low profile, low ESR, maximum CV	Maximum CV	Low ESR, maximum CV
TEMPERATURE RANGE	-55 °C to +125 °C (above 40 °C, voltage deratig is required)	-55 °C to +125 °C (above 85 °C, voltage derating is required)			
CAPACITANCE RANGE	330 µF to 2200 µF	1 µF to 2200 µF	1 µF to 1500 µF	0.1 µF to 1500 µF	1 µF to 1500 µF
VOLTAGE RANGE	6 V to 10 V	4 V to 50 V	4 V to 50 V	4 V to 50 V	4 V to 50 V
CAPACITANCE TOLERANCE	± 20 %	± 10 %, ± 20 %	± 10 %, ± 20 %	± 10 %, ± 20 %	± 10 %, ± 20 %
LEAKAGE CURRENT	0.01 CV or 0.5 µA, whichever is greater				
DISSIPATION FACTOR	14 % to 45 %	4 % to 50 %	4 % to 50 %	4 % to 20 %	4 % to 20 %
CASE CODES	C, M, X	S, A, B, C, D, R, M, X	A, B, C, D, R, M	T, S, A, B, C, D, G, M, R	B, C, D, R
TERMINATION	100 % matte tin	100 % matte tin standard, tin / lead and gold plated available			

SOLID TANTALUM CAPACITORS - CONFORMAL COATED

SERIES	597D	572D	595D	195D	194D
PRODUCT IMAGE					
TYPE	TANTAMOUNT™ chip, conformal coated				
FEATURES	Ultra low ESR, maximum CV, multi-anode	Low profile, maximum CV	Pad compatible with 194D and CWR06	US and European case sizes	Industrial version of CWR06 / CWR16
TEMPERATURE RANGE	-55 °C to +125 °C (above 85 °C, voltage derating is required)				
CAPACITANCE RANGE	10 µF to 1500 µF	2.2 µF to 220 µF	0.1 µF to 270 µF	0.1 µF to 330 µF	0.1 µF to 330 µF
VOLTAGE RANGE	4 V to 75 V	4 V to 35 V	4 V to 50 V	2 V to 50 V	4 V to 50 V
CAPACITANCE TOLERANCE	± 10 %, ± 20 %				
LEAKAGE CURRENT	0.01 CV or 0.5 µA, whichever is greater				
DISSIPATION FACTOR	6 % to 20 %	6 % to 26 %	4 % to 8 %	4 % to 8 %	4 % to 10 %
CASE CODES	V, D, E, R, F, Z, M, H	P, Q, S, A, B, T	A, B, D, E, F, G, H	C, S, V, X, Y, Z, R, A, B, D, E, F, G, H	A, B, C, D, E, F, G, H
TERMINATION	100 % matte tin standard, tin / lead solder plated available	100 % matte tin standard, gold plated available	100 % matte tin standard, tin / lead and gold plated available		Gold plated standard; tin / lead solder plated and hot solder dipped available

Revision: 12-Sep-17 3 Document Number: 40150

For technical questions, contact: tantalum@vishay.com



IXYS VUB145-16NOXT

Rectifier

Fig. 1 Forward current vs. voltage drop per diode

Fig. 2 Surge overload current vs. time per diode

Fig. 3 I^2t vs. time per diode

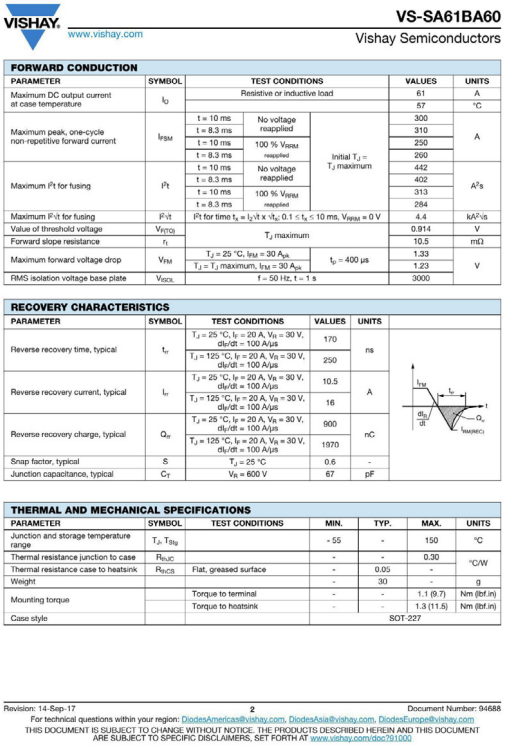
Fig. 4 Power dissipation vs. forward current and ambient temperature per diode

Fig. 5 Max. forward current vs. case temperature per diode

Fig. 6 Transient thermal impedance junction to case vs. time per diode

Constants for Z_{thJC} calculation:

i	R_{thi} (K/W)	t_i (s)
1	0.040	0.004
2	0.003	0.010
3	0.140	0.030
4	0.120	0.300
5	0.197	0.080



VISHAY. www.vishay.com

VS-SA61BA60

Vishay Semiconductors

FORWARD CONDUCTION

PARAMETER	SYMBOL	TEST CONDITIONS			VALUES	UNITS
Maximum DC output current at case temperature	I_O	Resistive or inductive load			61	A
					57	°C
Maximum peak, one-cycle non-repetitive forward current	I_{FSM}	t = 10 ms	No voltage reapplied	Initial T_J = T_J maximum	300	A
		t = 8.3 ms			310	
		t = 10 ms	100 % V_{RRM} reapplied		250	
		t = 8.3 ms			260	
Maximum I^2t for fusing	I^2t	t = 10 ms	No voltage reapplied		442	A^2s
		t = 8.3 ms			402	
		t = 10 ms	100 % V_{RRM} reapplied		313	
		t = 8.3 ms			284	
Maximum $I^2\sqrt{t}$ for fusing	$I^2\sqrt{t}$	I^2t for time $t_x = I_2\sqrt{t} \times \sqrt{t_x}$; $0.1 \le t_x \le 10$ ms, V_{RRM} = 0 V			4.4	$kA^2\sqrt{s}$
Value of threshold voltage	$V_{F(TO)}$	T_J maximum			0.914	V
Forward slope resistance	r_t				10.5	mΩ
Maximum forward voltage drop	V_{FM}	T_J = 25 °C, I_{FM} = 30 A_{pk}		t_p = 400 µs	1.33	V
		T_J = T_J maximum, I_{FM} = 30 A_{pk}			1.23	
RMS isolation voltage base plate	V_{ISOL}	f = 50 Hz, t = 1 s			3000	

RECOVERY CHARACTERISTICS

PARAMETER	SYMBOL	TEST CONDITIONS	VALUES	UNITS
Reverse recovery time, typical	t_{rr}	T_J = 25 °C, I_F = 20 A, V_R = 30 V, dI_F/dt = 100 A/µs	170	ns
		T_J = 125 °C, I_F = 20 A, V_R = 30 V, dI_F/dt = 100 A/µs	250	
Reverse recovery current, typical	I_{rr}	T_J = 25 °C, I_F = 20 A, V_R = 30 V, dI_F/dt = 100 A/µs	10.5	A
		T_J = 125 °C, I_F = 20 A, V_R = 30 V, dI_F/dt = 100 A/µs	16	
Reverse recovery charge, typical	Q_{rr}	T_J = 25 °C, I_F = 20 A, V_R = 30 V, dI_F/dt = 100 A/µs	900	nC
		T_J = 125 °C, I_F = 20 A, V_R = 30 V, dI_F/dt = 100 A/µs	1970	
Snap factor, typical	S	T_J = 25 °C	0.6	-
Junction capacitance, typical	C_T	V_R = 600 V	67	pF

THERMAL AND MECHANICAL SPECIFICATIONS

PARAMETER	SYMBOL	TEST CONDITIONS	MIN.	TYP.	MAX.	UNITS
Junction and storage temperature range	T_J, T_{Stg}		- 55	-	150	°C
Thermal resistance junction to case	R_{thJC}		-	-	0.90	°C/W
Thermal resistance case to heatsink	R_{thCS}	Flat, greased surface	-	0.05	-	
Weight			-	30	-	g
Mounting torque		Torque to terminal	-	-	1.1 (9.7)	Nm (lbf.in)
		Torque to heatsink	-	-	1.3 (11.5)	Nm (lbf.in)
Case style			SOT-227			

Revision: 14-Sep-17 2 Document Number: 94688

For technical questions within your region: DiodesAmericas@vishay.com, DiodesAsia@vishay.com, DiodesEurope@vishay.com

Fig. 2. Examples from the private dataset that show the difficulty and uniqueness of the data. Images include figures, table in tables, figures in tables and inconsistent structuring of rows and columns.

evaluating on the test set, it increases the recall of the proposed table detection network to 1.0.

- Unlike the previous research, the problem is split into two sub-problems: (1) Detection of tables on full documents, (2) Extrac-

tion of the detected tabular regions (cropping) and applying row–column detection on these crops. Solutions to both subproblems is constructed on top of the Mask-RCNN method augmented with transfer learning. Both solutions achieve state-of-the-art precision and recall.

The rest of the article is organized as follows: A review of the related work for table detection and table structure detection is given in Section 2. The network parameters and holistic system implementation are presented in detail in Section 3, followed by an evaluation and examination of their performance in Section 4. Finally, the article is concluded in Section 5, and future directions are given.

2. Related work and motivation

2.1. Motivation

There are millions of electronic component specification sheets available and these available set of datasheets changes regularly due to new product introduction and the obsolescence of existing products. This is a high-scalability problem which is extremely difficult to handle via manual efforts. As each manufacturer formats their documents differently, this defeats simpler automation approaches to this information extraction task required to achieve visibility to the global supply chain's available products. There have been significant efforts in this subject, because of the importance of this task. Existing solutions are based on image processing, statistical approaches, and certain methods rely on PDF metadata or text from the document, while others exploit machine and deep learning using images with no metadata.

2.2. PDF-based approaches

In this approach, character related checks and calculations, tab or space, and alignment checks are essential steps for the document processing. Fang et al. (2011) use PDF metadata to make use of visual separators such as white-space and dividing rules and geometric content layout information. They also attempt to improve the results by incorporating column detection into their table detection system. PDF-TREX (Oro and Ruffolo, 2009) is a heuristics-based approach to table detection. They rely on PDF metadata and build the table detections in a bottom-up fashion by combining text elements with alignment checks to form tabular regions.

2.3. Image-based approaches

This approach spans more data as it includes scanned and historical documents, thus it contains more challenges compared with other approaches. T-Recs (Kieninger and Dengel, 1998) is one of the earliest and most successful table detection methods that operate on images. T-Recs detects text blocks and then finds vertically adjacent text blocks to form columns. They depend on perfect detection of text blocks and single-line rows with perfect separation of columns. Tran et al. (2015) use Morphology method (Serra, 1983), which is designed to analyze and process geometrical structures. The authors apply morphological closing to connect adjacent text blocks to one another to form connected components. Then by the alignment and arrangement checks, they detect table regions. In Kara et al. (2019), the authors use the same idea but this time to detect rows and columns (i.e. table structure).

All the previously mentioned research build upon hand-crafted rules. However, using machine learning and deep learning is growing in popularity. Cesarini et al. (2002) convert documents to hierarchical MXY tree form. From this tree, they detect parallel horizontal and vertical lines and then create candidate table areas. Candidates are then verified using whitespaces or perpendicular lines. They use optimizations to create a definition for an optimal threshold to locate tables. However, there are other techniques used to solve tabular structure detection problem. Hao et al. (2016) designed weak hand-crafted rules to detect candidate tables and used deep learning to classify those regions as a table (or not a table). In addition to the cropped image representing those loosely extracted regions, they try feeding text from the PDF into the model as well but results are varying and do not improve significantly. Schreiber et al. (2017), utilize Faster-RCNN model to detect tables from document images. They use shallow backbone networks like VGG16 (Simonyan and Zisserman, 2014) for Faster-RCNN. But, they achieved state-of-the-art detection accuracy and precision on ICDAR 2013 dataset. They achieved this performance by stretching the images vertically to overcome some difficulties caused by the nature of the convolutional networks. This paper employs the same idea in the structure detection methods. Kavasidis et al. (2018) aim to solve table and chart segmentation problems with deep learning. They do not use a pre-trained network like most others but try to train their own model which is a modification of VGG16 and is a Fully Convolutional Network (FCN). The output of the model fed into Conditional Random Field (CRF) and binary classifiers to improve upon initial segmentations and confirm that the detections contain the correct object. Their training process is not trivial but they have achieved state-of-the-art detection results on the ICDAR 2013 test set as well. Traquair et al. (2019) presented a comparison between Faster-RCNN and RetinaNet (Lin et al., 2017b) and concluded that Faster-RCNN to be the best detection network for this task. This comparison is important because these are the two object detectors used on a variety of tasks and have different architectural designs. For example, RetinaNet is a single-stage detector with Feature Pyramid Network (FPN) (Lin et al., 2017a) architecture. Whereas, Faster-RCNN is a two-stage object detector with ROI-Pooling and there are no sophisticated design choices such as FPNs. These details will be further explained in Section 3.

Detecting tabular structures or charts is a necessary step to achieve a robust document information extraction tool. Table detection, in particular, has seen substantial research however, extraction of finer-grained table structures such as rows and columns have not received as much attention. One reason for this is, due to the nature of these structures, they tend to be narrow in one direction and tall on the other direction (Kara et al., 2019). For example, rows can have a height of one character and be as wideness as the entire page. Convolutional neural networks, have convolution filters (i.e. 7×7 filter on ResNet (He et al., 2016)) and pooling layers face a unique difficulty on tiny objects due to their large strides (Kara et al., 2019; Hu and Ramanan, 2017). Therefore, tabular structure detection is a difficult task for CNNs. CERMINE (Tkaczyk et al., 2015) is a metadata extraction method that uses machine learning to detect basic structures in documents such as email addresses or page number. It takes a PDF as input and then generates a hierarchical structure to represent the document, then extraction is done via machine learning. There is also more focused research done for information extraction from tables. Milosevic et al. (2019) focus on information extraction from biomedical documents with a combination of machine learning and rule-based methods. The first step is extracting tables then followed by structural processing and semantic operations. DeepDeSRT (Schreiber et al., 2017) uses an FCN-based model to recognize structures in the table crops after tables are detected with their table detection model. Their promising results with Faster-RCNN and FCN inspired us to combine these two techniques in the proposed approach. Previously (Kara et al., 2019) we proposed that Mask-RCNN is more effective than many other methods, including heuristics-based methods in the task of structure detection. Mask-RCNN is a more advanced version of Faster-RCNN with FCN and other additions (He et al., 2017). They show that Mask-RCNN method was the better choice for this task, at the expense of detection speed. But models in Kara et al. (2019) expects a perfect table extraction to reach those results. Not many works are published in this field but interested readers can read more about structure detection from Mao et al. (2003). To present the

results, Recall, Precision, Average Precision (AP, VOC 11-point average style) and F1 score which are standard metrics to compare different object detection methods (He et al., 2017; Lin et al., 2017b) are used. However, researchers also proposed specialized metrics for comparing the effectiveness of table detection methods as presented in Shahab et al. (2010). If one cares more about true-positives instead of false-negatives, these metrics can provide more insights on the detection performance.

2.4. Object detection

Images containing background pixels and tables are often visually salient, thus the tables can be separated from the rest of the document with visual cues. Using these characteristics, object detection has been selected as a path to automatically analyze a document. Convolutional neural networks have been shown to be state-of-the-art in the table detection task (Traquair et al., 2019) in terms of precision, recall and accuracy and probably even speed in some cases. Although CNNs are an essential part of the current state-of-the-art, there are important architectural enhancements to be made; backbone network to be used as a feature extractor, Feature Pyramid Networks (FPN) (Lin et al., 2017a) and Fully Convolutional Networks (FCN) (Shelhamer et al., 2017).

Faster-RCNN (Ren et al., 2015) is an object detector with a backbone network and Region Proposal Network (RPN). Backbone networks such as ResNet-50 extract features and then these features are fed into the RPN which then proposes candidate bounding boxes that may include an object of interest. Mask-RCNN (He et al., 2017) is almost identical to the Faster-RCNN but it employs additional methods to further improve the results. As presented in Ren et al. (2015), Faster-RCNN improved vastly upon previous object detection architectures and gained wide adoption (Traquair et al., 2019; Schreiber et al., 2017; Kavasidis et al., 2018).

Different backbone networks have various properties to their benefit. VGG16 (Simonyan and Zisserman, 2014) benefits from its speed and low parameter space whereas, ResNet-50 and ResNet-101 (He et al., 2016) networks are more capable of extracting features albeit at a reduced speed. Therefore, diverse detection tasks benefit from these networks. They are being used in many different object detection methods including RetinaNet (Lin et al., 2017b) and Faster-RCNN (Ren et al., 2015). Due to the ResNet networks' depth, it can encode a large amount of data and extract rich features. AlexNet (Krizhevsky et al., 2012) showed that in a deep convolutional neural network, the first layers encode low-level information such as edges and corners and color related information, where later layers encode more high-level information that is representative of objects. Traditionally, features are extracted from one of the later convolutional layers from the backbone network. However, the convolution multiplications causes low-level information to be lost. Hence, for problems that require these low-level details in addition to the high-level abstract features, feature pyramid networks are useful.

The feature pyramid network (Lin et al., 2017a) architecture borrows the idea of using different scales of features from conventional image processing. Instead of scaling the image and extracting features from different levels, FPN extracts features from different levels of a CNN in a bottom-up (high-resolution to low-resolution) fashion and combines them in a top-down (low-resolution to high-resolution) fashion. In table images, rows and columns are small in size. Hence, after fifty layers of convolution and pooling operations, there are very few meaningful features which remain. Therefore, table structure detection benefits from FPN.

Shelhamer et al. (2017) proposed the Fully Convolutional Networks (FCN) and it is used to segment objects at the pixel-level instead of via chunky bounding box detection. Since the operations are done with the objective of pixel-level detection, models with the FCN have to retain and track information related to every pixel, if possible. Table structure detection can be formulated as a segmentation task to capture small rows and columns. In the previous work, the study tests this idea and shows that Mask-RCNN outperforms Faster-RCNN (Kara et al., 2019). The primary differences between the two networks can be simplified to Mask-RCNN having both FPN and FCN on top of the Faster-RCNN method.

2.5. Mask-RCNN

For both sub-tasks of table detection and table structure detection, this work proposes a Mask-RCNN based method to benefit from the aforementioned architectures. Mask-RCNN (He et al., 2017) builds upon Faster-RCNN (Ren et al., 2015) method. Faster-RCNN consists of two stages; the first stage is called Region Proposal Network (RPN) which is used to propose possible object bounding boxes and the second stage is adopted from earlier Fast-RCNN (Girshick, 2015) model and called RoIPool. RoIPool takes the output of RPN and extracts features to form a fixed-size feature map. The extracted feature map is then sent to the regression and classification branches to finish the regression of the box location and the classification of the box. Therefore, Faster-RCNN has two outputs; object classes and boxes. Box prediction and classification of the boxes are performed in parallel. Therefore, classes are not competing while generating boxes. Mask-RCNN adopts the first stage (RPN) as is. In the second stage, an FCN branch is added to generate binary masks that predict a mask for each fixed-size feature map. Generating binary masks mean the classes are not competing in the mask prediction as well. All these calculations and predictions are done in parallel within the branches. The overall flow of the Mask-RCNN method is shown in Fig. 3.

Mask-RCNN uses a multi-task loss on each of the RoIs extracted. The loss is defined as; $L = L_{mask} + L_{class} + L_{box}$. The bounding box loss L_{box} is *smooth-L1* loss whereas the classification loss L_{cls} is log-loss for true classes. These loss functions are adopted directly from Fast-RCNN and more details can be found in Girshick (2015). The mask loss L_{mask} is the average binary cross-entropy loss. The mask branch outputs several masks with dimensions $m \times m$ (in this case, $m = 28$). To these masks, Mask-RCNN applies a per-pixel sigmoid loss function to calculate final L_{mask}. This formulation of L_{mask} leads to the generation of masks without classes, as in other branches of Faster-RCNN. The classification of the masks relies on the classification branch as well. RoIPool in Faster-RCNN performs the extraction of features from RoIs with a method called *quantization*. But the process introduces misalignment between the extracted features and the corresponding RoI. While this is not crucial in object detection, in a pixel-level detection task this is vital. Therefore, Mask-RCNN introduces RoIAlign and switches to using binary interpolation of RoI features to extract fixed-sized feature maps. More detailed explanations can be found in He et al. (2017).

Batch Normalization (BN) (Ioffe and Szegedy, 2015) is introduced to increase training performance by solving the internal covariate shift problem. That is, it works as a regularizer but operates between the layers of deep networks. Batch size of 1 render BN useless due to it working by averaging the mini-batches (Ioffe and Szegedy, 2015). Therefore, Group Normalization (GN) (Wu and He, 2018) is used. Yuxin et al. proposed GN to overcome this issue with BN. Unlike BN, GN divides the channels inputted to a layer by a fixed number (in this case *group number* = 32) and normalizes the groups. Wu and He (2018) shows that with small batch sizes, group normalization increases the precision over batch normalization.

3. Methodology

The proposed holistic solution, also referred to as an E2E solution, is based on the information discovery from tabular regions which can be split into two main sub-tasks; table detection and structure detection.

For the E2E system, four types of solutions are implemented; a pure E2E system, a vanilla system, a judge based system, and an agreement-based model. Other than the purely E2E method, each method consists

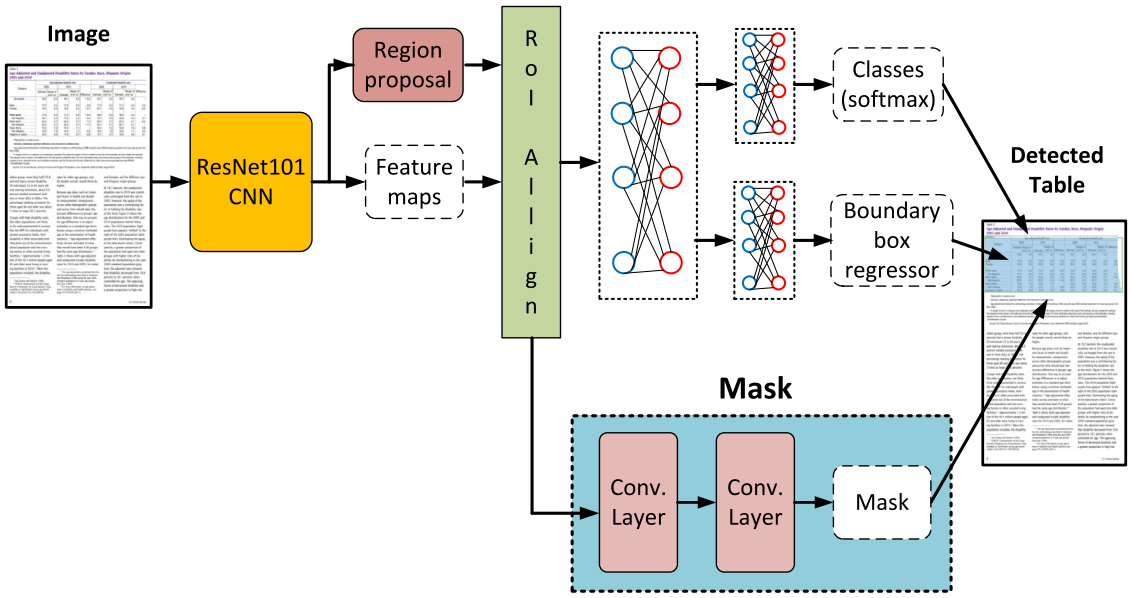


Fig. 3. Mask-RCNN architecture integrated with the table detection problem.

of multiple discrete components. Judge and agreement models are developed to address the problems in table detection accuracy to confidently feed tables into the structure detection method.

3.1. Pure E2E system

The pure E2E system consists of a single Mask-RCNN based method that is trained to detect rows and columns on document page images. The problems encountered in this system were that rows and columns encompass only a few pixels in height or width and therefore, rows were obscured by the information loss. To solve this issue, the images are stretched to ensure a minimum of 50 pixels along their shortest side. Stretching gave us close to 20% increase in recall and precision. The pure E2E model is not affected by the inconsistent detections of other models which result from passing table detections to structure detection. It is a single-pass method and therefore the fastest of the considered methods. Additionally, the model was trained to detect rows and columns and tables as well which increased the average precision.

3.2. Vanilla two-stage system

The vanilla E2E system forms the simplest proposal of th two-stage methods, as it consists of two Mask-RCNN (He et al., 2017) models operating sequentially, with the output of the first fed as the input of the second. These two models operate without any improvements, each trained for their specific task (of table detection, and structure detection). As can be seen in Fig. 6a, after converting the desired PDF into page images, the images are then fed into the table detection network. Table detections are then cropped and served into the structure detection network. Several example detections (both correct and erroneous) can be found in Figs. 4–5. The figures show some interesting cases where multiple tables are in a page or where there are nested tables inside each other. The false-positive cases are presented where the model detected a table but there was no table. Different post-processing methods were experimented with to eliminate false-positives and they will be presented in the next sub-section.

The vanilla system acts as a baseline approach derived directly from previous work done for both table extraction, and structure detection. This method is the simplest and requires the fewest calculations, meaning it is also the fastest two-stage method to run. Concerns for the performance of this method arise from the propagation of errors from network to network. False-positives or incomplete detection from table detections will affect structure detection performance. Since no validation of the table detections exists in this method, further fine-tuning of the table detection network was done to attain optimal recall and precision.

3.3. Improvements to two-stage method

Since false-positive detections are problematic in a fully autonomous workflow, the aim is to further improve the detection performance and eliminate false-positives if possible. Therefore, experiments were performed with a couple different methods to augment the baseline two-stage method.

3.3.1. Judging approach

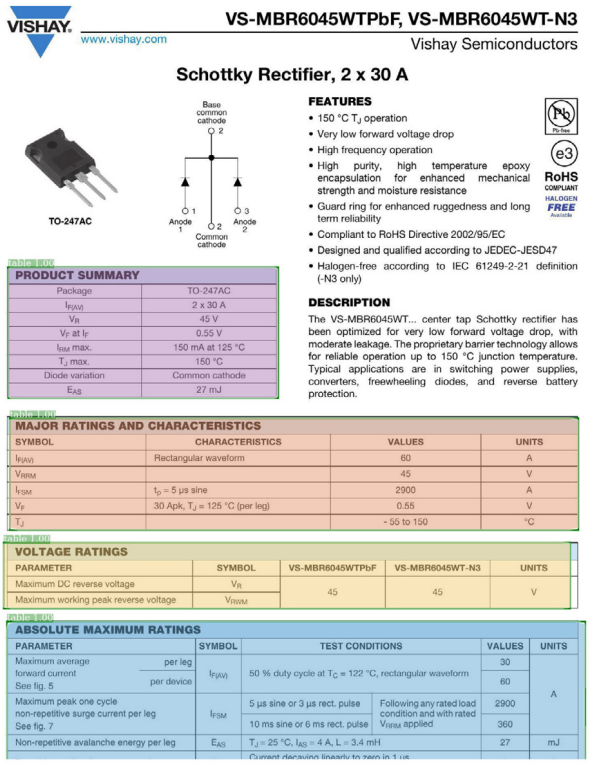
As mentioned for the vanilla system, there is a need to validate table detection results before detecting structures. The judge system attempts to address the reliability of the table detections being fed into the structure detections. As seen in Fig. 6b, where the PDF images are fed purely into the table detection network in the vanilla system, they are instead fed into two networks to be fed into a judging mechanism. The first network is the same as before, a table detection network, and the other is the pure E2E structure detection network which instead of being trained exclusively on table images, is trained on the whole page images. The output of the structure detection network contains row and column predictions. Table detection is performed as usual. After both detections, an output similar to that is shown in Figs. 7a and 7b. When both networks have performed detections on the whole page image, the judge algorithm compares the overlapping segments.

$$IoU = \frac{Area\ of\ Intersection}{Area\ of\ Union} \tag{1}$$

Overlaps are calculated in terms of Intersection over Union (IoU) as it is used to accept or reject bounding boxes in object detection methods and the calculation is given in Eq. (1).

$$Judge\ Approve_i = \begin{cases} True, & IoU_{i,j} > 1 - Conf_i \\ False, & \text{otherwise} \end{cases} \tag{2}$$

The judge looks for an IoU greater than the preselected confidence score threshold of the system as shown in Eq. (2). The threshold defined as $1 - Conf_i$. Where i is ith detected table in the image and where j is jth detected internal structure in the image and $Conf_i$ is the confidence of the model that ith detection is a table. If this condition is satisfied,



VISHAY www.vishay.com

VS-MBR6045WTPbF, VS-MBR6045WT-N3

Vishay Semiconductors

Schottky Rectifier, 2 x 30 A

TO-247AC

Base common cathode 2 — Anode 1 — Common cathode 2 — Anode 2

FEATURES

- 150 °C T_J operation
- Very low forward voltage drop
- High frequency operation
- High purity, high temperature epoxy encapsulation for enhanced mechanical strength and moisture resistance
- Guard ring for enhanced ruggedness and long term reliability
- Compliant to RoHS Directive 2002/95/EC
- Designed and qualified according to JEDEC-JESD47
- Halogen-free according to IEC 61249-2-21 definition (-N3 only)

DESCRIPTION

The VS-MBR6045WT... center tap Schottky rectifier has been optimized for very low forward voltage drop, with moderate leakage. The proprietary barrier technology allows for reliable operation up to 150 °C junction temperature. Typical applications are in switching power supplies, converters, freewheeling diodes, and reverse battery protection.

PRODUCT SUMMARY

Package	TO-247AC
$I_{F(AV)}$	2 x 30 A
V_R	45 V
V_F at I_F	0.55 V
I_{RM} max.	150 mA at 125 °C
T_J max.	150 °C
Diode variation	Common cathode
E_{AS}	27 mJ

MAJOR RATINGS AND CHARACTERISTICS

SYMBOL	CHARACTERISTICS	VALUES	UNITS
$I_{F(AV)}$	Rectangular waveform	60	A
V_{RRM}		45	V
I_{FSM}	t_p = 5 µs sine	2900	A
V_F	30 Apk, T_J = 125 °C (per leg)	0.55	V
T_J		- 55 to 150	°C

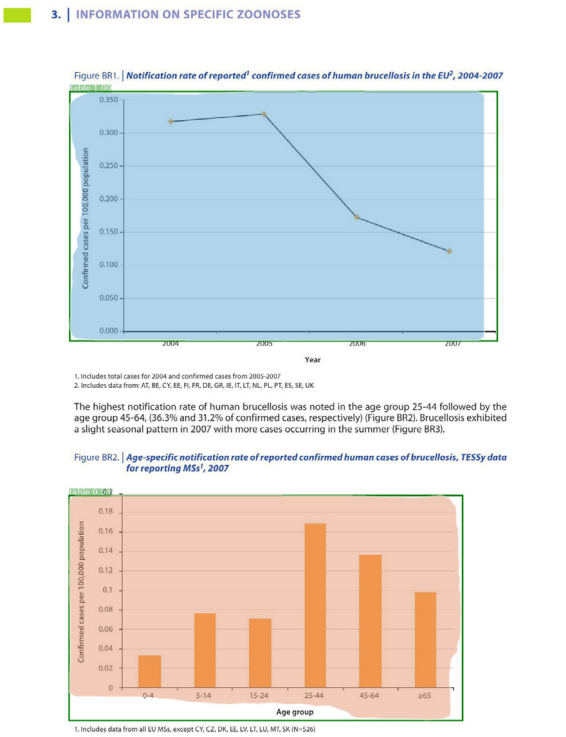
VOLTAGE RATINGS

PARAMETER	SYMBOL	VS-MBR6045WTPbF	VS-MBR6045WT-N3	UNITS
Maximum DC reverse voltage	V_R	45	45	V
Maximum working peak reverse voltage	V_{RWM}			

ABSOLUTE MAXIMUM RATINGS

PARAMETER		SYMBOL	TEST CONDITIONS		VALUES	UNITS
Maximum average forward current See fig. 5	per leg	$I_{F(AV)}$	50 % duty cycle at T_C = 122 °C, rectangular waveform		30	A
	per device				60	
Maximum peak one cycle non-repetitive surge current per leg See fig. 7		I_{FSM}	5 µs sine or 3 µs rect. pulse	Following any rated load condition and with rated V_{RRM} applied	2900	
			10 ms sine or 6 ms rect. pulse		360	
Non-repetitive avalanche energy per leg		E_{AS}	T_J = 25 °C, I_{AS} = 4 A, L = 3.4 mH		27	mJ
			Current decaying linearly to zero in 1 µs			

(a) Correct table detections



3. | INFORMATION ON SPECIFIC ZOONOSES

Figure BR1. | *Notification rate of reported[1] confirmed cases of human brucellosis in the EU[2], 2004-2007*

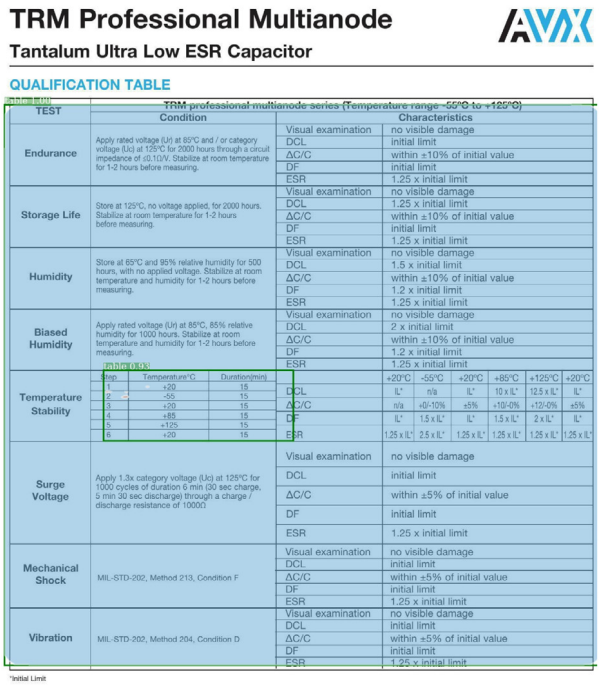
1. Includes total cases for 2004 and confirmed cases from 2005-2007
2. Includes data from: AT, BE, CY, EE, FI, FR, DE, GR, IE, IT, LT, NL, PL, PT, ES, SE, UK

The highest notification rate of human brucellosis was noted in the age group 25-44 followed by the age group 45-64, (36.3% and 31.2% of confirmed cases, respectively) (Figure BR2). Brucellosis exhibited a slight seasonal pattern in 2007 with more cases occurring in the summer (Figure BR3).

Figure BR2. | *Age-specific notification rate of reported confirmed human cases of brucellosis, TESSy data for reporting MSs[1], 2007*

1. Includes data from all EU MSs, except CY, CZ, DK, EE, LV, LT, LU, MT, SK (N=526)

(a) False-positives - Figures detected as tables



TRM Professional Multianode

Tantalum Ultra Low ESR Capacitor

AVX

QUALIFICATION TABLE

TRM professional multianode series (Temperature range -55°C to +125°C)

TEST	Condition	Characteristics	
Endurance	Apply rated voltage (Ur) at 85°C and / or category voltage (Uc) at 125°C for 2000 hours through a circuit impedance of ≤0.1Ω/V. Stabilize at room temperature for 1-2 hours before measuring.	Visual examination	no visible damage
		DCL	initial limit
		ΔC/C	within ±10% of initial value
		DF	initial limit
		ESR	1.25 x initial limit
Storage Life	Store at 125°C, no voltage applied, for 2000 hours. Stabilize at room temperature for 1-2 hours before measuring.	Visual examination	no visible damage
		DCL	1.25 x initial limit
		ΔC/C	within ±10% of initial value
		DF	initial limit
		ESR	1.25 x initial limit
Humidity	Store at 85°C and 95% relative humidity for 500 hours, with no applied voltage. Stabilize at room temperature and humidity for 1-2 hours before measuring.	Visual examination	no visible damage
		DCL	1.5 x initial limit
		ΔC/C	within ±10% of initial value
		DF	1.2 x initial limit
		ESR	1.25 x initial limit
Biased Humidity	Apply rated voltage (Ur) at 85°C, 85% relative humidity for 1000 hours. Stabilize at room temperature and humidity for 1-2 hours before measuring.	Visual examination	no visible damage
		DCL	2 x initial limit
		ΔC/C	within ±10% of initial value
		DF	1.2 x initial limit
		ESR	1.25 x initial limit
Temperature Stability	Step / Temperature °C / Duration (min): 1 +20 15; 2 -55 15; 3 +20 15; 4 +85 15; 5 +125 15; 6 +20 15		+20°C / -55°C / +20°C / +85°C / +125°C / +20°C
		DCL	IL* / n/a / IL* / 10 x IL* / 12.5 x IL* / IL*
		ΔC/C	n/a / +0/-10% / ±5% / +10/-0% / +12/-0% / ±5%
		DF	IL* / 1.5 x IL* / IL* / 1.5 x IL* / 2 x IL* / IL*
		ESR	1.25 x IL* / 2.5 x IL* / 1.25 x IL* / 1.25 x IL* / 1.25 x IL* / 1.25 x IL*
Surge Voltage	Apply 1.3x category voltage (Uc) at 125°C for 1000 cycles of duration 5 min (30 sec charge, 5 min 30 sec discharge) through a charge / discharge resistance of 1000Ω	Visual examination	no visible damage
		DCL	initial limit
		ΔC/C	within ±5% of initial value
		DF	initial limit
		ESR	1.25 x initial limit
Mechanical Shock	MIL-STD-202, Method 213, Condition F	Visual examination	no visible damage
		DCL	initial limit
		ΔC/C	within ±5% of initial value
		DF	initial limit
		ESR	1.25 x initial limit
Vibration	MIL-STD-202, Method 204, Condition D	Visual examination	no visible damage
		DCL	initial limit
		ΔC/C	within ±5% of initial value
		DF	initial limit
		ESR	1.25 x initial limit

*Initial Limit

(b) A table inside another table is detected

Fig. 4. Correct table detection examples: True-positive: The detected area contains at least a table.

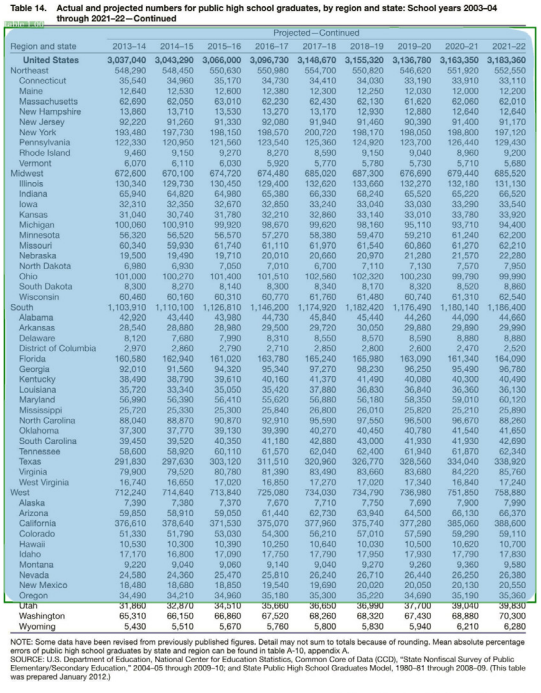


Table 14. **Actual and projected numbers for public high school graduates, by region and state: School years 2003–04 through 2021–22—Continued**

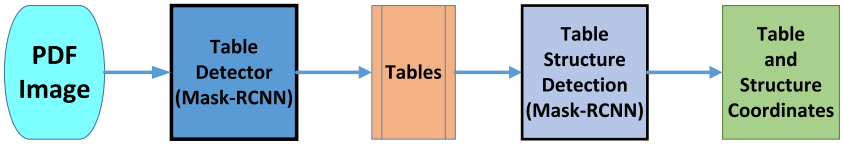
Region and state	Projected—Continued								
	2013–14	2014–15	2015–16	2016–17	2017–18	2018–19	2019–20	2020–21	2021–22
United States	**3,037,040**	**3,043,290**	**3,066,000**	**3,096,730**	**3,148,670**	**3,155,320**	**3,136,780**	**3,163,350**	**3,183,360**
Northeast	548,290	548,450	550,630	550,980	554,700	550,820	546,620	551,920	552,550
Connecticut	35,540	34,960	35,170	34,730	34,410	34,030	33,190	33,910	33,110
Maine	12,640	12,530	12,600	12,380	12,300	12,250	12,030	12,000	12,200
Massachusetts	62,690	62,050	63,010	62,230	62,430	62,130	61,620	62,060	62,010
New Hampshire	13,860	13,710	13,530	13,270	13,170	12,930	12,880	12,640	12,640
New Jersey	92,220	91,260	91,330	92,080	91,940	91,460	90,390	91,400	91,170
New York	193,480	197,730	198,150	198,570	200,720	198,170	198,050	198,800	197,120
Pennsylvania	122,330	120,950	121,560	123,540	125,360	124,920	123,700	126,440	129,430
Rhode Island	9,460	9,150	9,270	8,270	8,590	9,150	9,040	8,960	9,200
Vermont	6,070	6,110	6,030	5,920	5,770	5,780	5,730	5,710	5,680
Midwest	672,600	670,100	674,720	674,480	685,020	687,300	676,690	679,440	685,520
Illinois	130,340	129,730	130,450	129,400	132,620	133,660	132,270	132,180	131,130
Indiana	65,940	64,820	64,980	65,380	66,330	68,240	65,520	65,220	66,520
Iowa	32,310	32,350	32,670	32,850	33,240	33,040	33,030	33,290	33,540
Kansas	31,040	30,740	31,780	32,210	32,860	33,140	33,010	33,780	33,920
Michigan	100,060	100,910	99,920	98,670	99,620	98,160	95,110	93,710	94,400
Minnesota	56,320	56,520	56,570	57,270	58,380	59,470	59,210	61,240	62,200
Missouri	60,340	59,930	61,740	61,110	61,970	61,540	60,860	61,270	62,210
Nebraska	19,500	19,490	19,710	20,010	20,660	20,970	21,280	21,570	22,280
North Dakota	6,980	6,930	7,050	7,010	6,700	7,110	7,130	7,570	7,950
Ohio	101,000	100,270	101,400	101,510	102,560	102,320	100,230	99,790	99,990
South Dakota	8,300	8,270	8,140	8,300	8,340	8,170	8,320	8,520	8,860
Wisconsin	60,460	60,160	60,310	60,770	61,760	61,480	60,740	61,310	62,540
South	1,103,910	1,110,100	1,126,810	1,146,200	1,174,920	1,182,420	1,176,490	1,180,140	1,186,400
Alabama	42,920	43,440	43,980	44,730	45,840	45,440	44,260	44,090	44,660
Arkansas	28,540	28,880	28,980	29,500	29,720	30,050	29,880	29,890	29,960
Delaware	8,120	7,680	7,990	8,310	8,550	8,570	8,590	8,880	8,880
District of Columbia	2,970	2,860	2,790	2,710	2,850	2,800	2,800	2,470	2,520
Florida	160,580	162,940	161,020	163,780	165,240	165,980	163,090	161,340	164,090
Georgia	92,010	91,560	94,320	95,340	97,270	98,230	96,250	95,490	96,780
Kentucky	38,490	38,790	39,610	40,160	41,370	41,490	40,080	40,300	40,490
Louisiana	35,720	33,340	35,050	35,420	37,880	36,830	36,840	36,360	36,130
Maryland	56,990	56,390	56,410	55,620	56,880	56,180	58,350	59,010	60,120
Mississippi	25,720	25,330	25,300	25,840	26,800	26,010	25,820	25,210	25,890
North Carolina	88,040	88,870	90,870	92,910	95,590	97,550	96,500	96,670	88,260
Oklahoma	37,300	37,770	39,130	39,390	40,270	40,450	40,780	41,540	41,650
South Carolina	39,450	39,520	40,350	41,180	42,880	43,000	41,930	41,930	42,690
Tennessee	58,600	58,920	60,110	61,570	62,040	62,400	61,940	61,870	62,340
Texas	291,830	297,630	303,120	311,510	320,960	326,770	328,560	334,040	338,920
Virginia	79,900	79,520	80,780	81,390	83,490	83,660	83,680	84,220	85,760
West Virginia	16,740	16,650	17,020	16,850	17,270	17,020	17,340	16,840	17,240
West	712,240	714,640	713,840	725,080	734,030	734,790	736,980	751,850	758,880
Alaska	7,390	7,380	7,370	7,670	7,710	7,750	7,690	7,900	7,990
Arizona	59,850	58,910	59,050	61,440	62,730	63,940	64,500	66,130	66,370
California	376,610	378,640	371,530	375,070	377,960	375,740	377,280	385,060	388,600
Colorado	51,330	51,790	53,030	54,300	56,210	57,010	57,590	59,290	59,110
Hawaii	10,530	10,300	10,390	10,250	10,840	10,030	10,500	10,820	10,700
Idaho	17,170	16,800	17,090	17,750	17,790	17,950	17,930	17,790	17,830
Montana	9,220	9,040	9,060	9,140	9,040	9,270	9,260	9,380	9,580
Nevada	24,580	24,360	25,470	25,810	26,240	26,710	26,440	26,250	26,380
New Mexico	18,480	18,680	18,850	19,540	19,690	20,020	20,050	20,130	20,550
Oregon	34,490	34,210	34,960	35,180	35,300	35,220	34,690	35,190	35,360
Utah	31,860	32,870	34,510	35,660	36,650	36,990	37,700	39,040	39,830
Washington	65,310	66,150	66,860	67,520	68,260	68,320	67,430	68,880	70,300
Wyoming	5,430	5,510	5,670	5,760	5,800	5,830	5,940	6,210	6,280

NOTE: Some data have been revised from previously published figures. Detail may not sum to totals because of rounding. Mean absolute percentage errors of public high school graduates by state and region can be found in table A-10, appendix A.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Common Core of Data (CCD), "State Nonfiscal Survey of Public Elementary/Secondary Education," 2004–05 through 2009–10; and State Public High School Graduates Model, 1980–81 through 2008–09. (This table was prepared January 2012.)

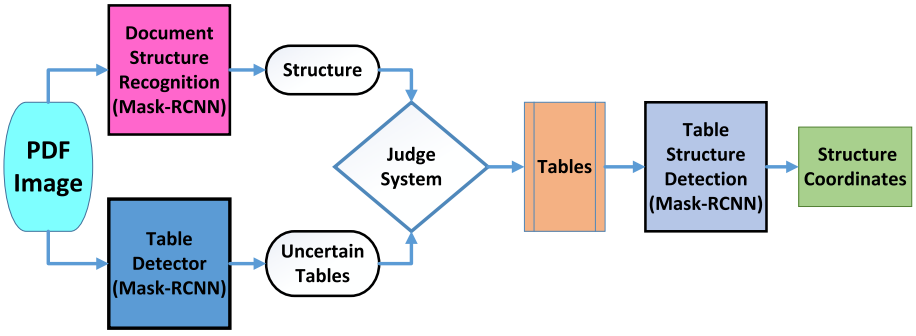
(b) Detector failed to detect some parts of the table

Fig. 5. Incorrect table detection examples including false-positive and partially detected table position. False-positive: The detected area does not contain any table.

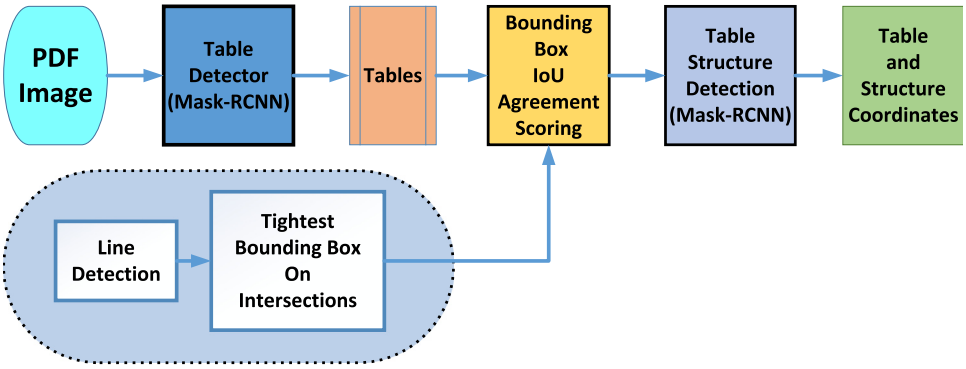
then only those segment judged to be a valid table will continue to the second phase; structure detection. Inefficiencies in the judging system arise from the large computation time taken per page, due to the use of 3 networks to perform a detection per page. The judging system is essentially used to decrease the number of false-positives.



(a) Vanilla System Flow Diagram



(b) Judge System Flow Diagram



(c) Agreement Model Flow Diagram

Fig. 6. System flow diagrams for proposed approach.

3.3.2. Agreement based approach

The E2E system is developed to attain optimal table detections to feed into the table structure detection component. The agreement based model is implemented using a heuristic based approach to address the merging and expansion of table bounding boxes. The agreement algorithm handles the detected tables with their confidences and evaluates the agreement scoring. Based on observations, the table detection network occasionally generates overlapping table detections with similar confidence levels. If the IoU between any two tables is between 0.8 and 0.5 and the confidence of both of the tables are less then 80%, no change is performed. If the IoU between any two tables is higher than 0.8 and the smaller table has a higher confidence score, the smaller table is expanded with the ratio $1-(Conf_i-Conf_j)$ where i is the smaller table and j is the larger table. This formula decides that if the smaller table is 10% more confident than the larger table, the small table will be expanded up to 90% of the difference between the tables. As can be seen in Fig. 8, the bounding box is always expanded either vertically or horizontally, depending on the functions of confidence and overlap, addressing the issue of partial detections of tables.

An enhancement to the agreement system aims to expand boxes by taking advantage of bounding lines that surround many tables. The overall flow of this algorithm can be seen in Fig. 9. Though not always associated with tabular structures, tables often contain bounding and separating lines, which can be used to highlight areas of interest and can be exploited by table detection models. The intersection of these lines can be used to define the minimum and maximum points of likely tables. In Figs. 7c and 7d an example case is presented where line agreement model increases the accuracy of the table detection. Since no false-positives are being eliminated, the precision metric will be the same. But, the two-stage holistic system metrics rely on how accurate table detections are, therefore it is important to increase the accuracy of tables. The design of this system is such that the modularity allows for the inputting of any type of line detection algorithm, as well as any other bounding boxes to act as further refinements. Here, Canny Edge Detection (Canny, 1987) is used with threshold values of 150 and 350 and an aperture size of 5. Following the edge detection, Probabilistic Hough Transform (Kiryati et al., 1991) with the Canny edge detection image was used as input and parameters:rho of 1, theta of $\frac{\pi}{180}$, a threshold of 100, a minimum length of 400, and finally a maximum line gap of 5 pixels. The objective of these parameters being to find only straight, continuous lines as it is not expected to find all possible positions of tables, it aims to highlight the complete bounds of those that are found. Following the discovery of lines, intersecting lines are extracted and a tight bounding box that encompasses all the intersecting lines is found. Then, further check is performed to see if this line has an IoU of higher than 0.8 with any table. If so, the table is extended to cover this tight line bounding box.

TABLE IV

^{123}I MEASURED IN J-13/HBR SOLUTION SAMPLES

Test	Days	^{123}I (pCi/ml)*	$\pm 10^{-5}$ Inventory**
Bare Fuel	63	0.52	5.6
	223	0.72	7.5
Slit Defect	63	0.29	3.0
	223	0.38	4.0
Hole Defects	63	0.0019	0.020
	223	0.0050	0.053
Undefected	63	0.0060	0.064
	223	0.0054	0.057

*Average for 2 replicate samples measured by neutron activation analysis.
**(pCi/ml)(250 ml)/(10^{-5} of specimen inventory).

TABLE V

H. B. ROBINSON FUEL IN J-13 WATER
URANIUM RELEASE DATA (μg)

	Bare Fuel	Slit Defect	Hole Defects	Undefected
∑ Solution Samples	211	4.3	0.50	1.26
∑ Rod Samples	34	0.3	<0.17	<0.19
Final Solution [U (μg/ml)]*	300 (1.2)	23.8 (0.095)	1.50 (0.006)	3.25 (0.013)
Strip	2700	1.5	0.60	0.60
Rinse**	550	1.8	0.60	0.27
∑ Above	3795	31.7	<3.37	<5.57
Divided by 10^{-5} Inv.	5.42	0.044	<0.0047	<0.008

*Unfiltered 223-day final solution uranium content in μg/ml given in parentheses.
**Bare fuel rinse solution was 0.4 μm filtered before analysis.

14

(a) Judge Module - E2E Structure Detection

TABLE IV

^{123}I MEASURED IN J-13/HBR SOLUTION SAMPLES

*Average for 2 replicate samples measured by neutron activation analysis.
**(pCi/ml)(250 ml)/(10^{-5} of specimen inventory).

TABLE V

H. B. ROBINSON FUEL IN J-13 WATER
URANIUM RELEASE DATA (μg)

*Unfiltered 223-day final solution uranium content in μg/ml given in parentheses.
**Bare fuel rinse solution was 0.4 μm filtered before analysis.

14

(b) Judge Module - Table Detection

Table 2.
Age-Adjusted and Unadjusted Disability Rates by Gender, Race, Hispanic Origin: 2005 and 2010

Category	Age-adjusted disability rate					Unadjusted disability rate				
	2005		2010			2005		2010		
	Estimate	Margin of error (±)	Estimate	Margin of error (±)	Difference	Estimate	Margin of error (±)	Estimate	Margin of error (±)	Difference
All people	18.6	0.3	18.1	0.3	*−0.5	18.7	0.3	18.7	0.3	–
Male	17.9	0.4	17.6	0.4	−0.3	17.3	0.4	17.4	0.4	0.2
Female	19.0	0.3	18.3	0.4	*−0.7	20.1	0.3	19.8	0.4	−0.2
White alone	17.9	0.3	17.4	0.3	*−0.5	18.6	0.3	18.5	0.3	–
Not Hispanic	18.1	0.4	17.6	0.4	−0.4	19.7	0.4	19.8	0.4	0.1
Black alone	23.2	0.7	22.2	0.7	−1.0	20.4	0.7	20.3	0.7	−0.2
Not Hispanic	23.3	0.7	22.3	0.7	*−1.0	20.7	0.7	20.7	0.7	–
Asian Alone	14.5	1.3	14.5	1.1	–	12.4	1.2	13.0	1.0	0.6
Not Hispanic	14.6	1.3	14.4	1.1	−0.2	12.5	1.2	13.0	1.1	0.5
Hispanic or Latino	18.4	0.9	17.8	0.7	−0.6	13.1	0.7	13.2	0.6	0.1

– Represents or rounds to zero.
* Denotes a statistically significant difference at the 90 percent confidence level.
[1] Age-adjustments followed the methodology described in Anderson and Rosenberg (1998) using the year 2000 standard population by 5-year age groups from Day (1996).
[2] A margin of error is a measure of an estimate's variability. The larger the margin of error in relation to the size of the estimate, the less reliable the estimate. The margins of error shown in this table are for the 90 percent confidence level. For more information about the source and accuracy of the estimates, including margins of error, standard errors, and confidence intervals, see the Source and Accuracy Statement at <http://www.census.gov/sipp/sourceacc/S&A08_W1toW6(S&A-13).pdf>.
Source: U.S. Census Bureau, Survey of Income and Program Participation, June–September 2005 and May–August 2010.

oldest group, more than half (55.8 percent) had a severe disability. Of individuals 55 to 64 years old and nearing retirement, about 6.0 percent needed assistance with one or more ADLs or IADLs. The percentage needing assistance for those aged 80 and older was about 5 times as large (30.2 percent).

rates for older age groups, and for people overall, would likely be higher.

Because age plays such an important factor in health and disability measurement, comparisons across other demographic groups and across time should take into

and females, and for different race and Hispanic origin groups.

At 18.7 percent, the unadjusted disability rate in 2010 was statistically unchanged from the rate in 2005, however, the aging of the population was a contributing fac-

(c) Agreement Module - Line Bounding Box

Table 2.
Age-Adjusted and Unadjusted Disability Rates by Gender, Race, Hispanic Origin: 2005 and 2010

– Represents or rounds to zero.
* Denotes a statistically significant difference at the 90 percent confidence level.
[1] Age-adjustments followed the methodology described in Anderson and Rosenberg (1998) using the year 2000 standard population by 5-year age groups from Day (1996).
[2] A margin of error is a measure of an estimate's variability. The larger the margin of error in relation to the size of the estimate, the less reliable the estimate. The margins of error shown in this table are for the 90 percent confidence level. For more information about the source and accuracy of the estimates, including margins of error, standard errors, and confidence intervals, see the Source and Accuracy Statement at <http://www.census.gov/sipp/sourceacc/S&A08_W1toW6(S&A-13).pdf>.
Source: U.S. Census Bureau, Survey of Income and Program Participation, June–September 2005 and May–August 2010.

oldest group, more than half (55.8 percent) had a severe disability. Of individuals 55 to 64 years old and nearing retirement, about 6.0 percent needed assistance with one or more ADLs or IADLs. The percentage needing assistance for those aged 80 and older was about

rates for older age groups, and for people overall, would likely be higher.

Because age plays such an important factor in health and disability measurement, comparisons across other demographic groups

and females, and for different race and Hispanic origin groups.

At 18.7 percent, the unadjusted disability rate in 2010 was statistically unchanged from the rate in 2005, however, the aging of the population was a contributing fac-

(d) Agreement Module - Table Detection

Fig. 7. Example cases for post-processing steps. Different colors stand for segmentation of different instances of the same class.

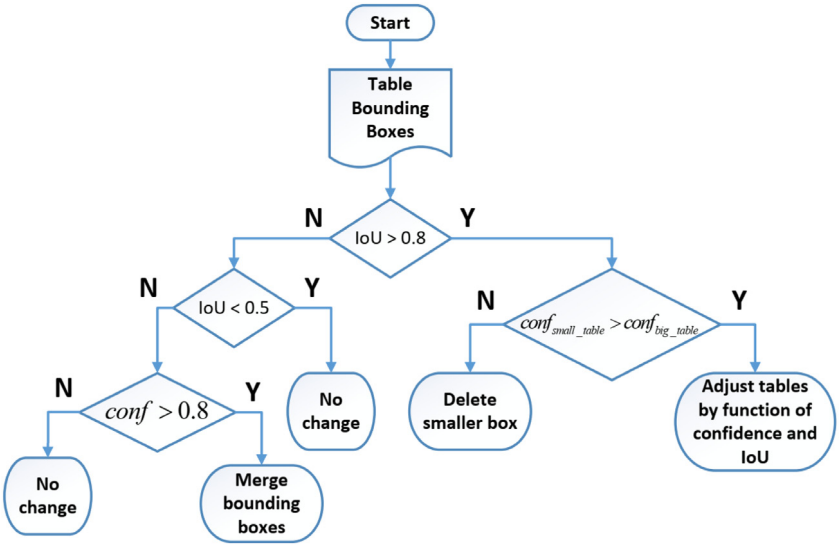


Fig. 8. Agreement module — table agreement algorithm flow.

Table 1
Datasets — Train and test splits.

Dataset	Total images	Training and validation	Test
ICDAR 2013	238	–	238
ICDAR 2017	2417	2217	200
Marmot	2000	829	–
UNLV	424	350	50
Private	2323	2123	200

3.4. Dataset

The datasets used consist of the publicly available ICDAR 2013 (238 images) (Karatzas et al., 2013), ICDAR 2017 Page Object Detection (POD) dataset (2417 images),[1] UNLV (424 images) (Rice et al., 2012), Marmot (2000 images)[2] as well as a private dataset of 2323 images aggregated by Lytica and hand labeled by us, that consists of the private training set of 2123 images and the private test set of 200 images. Across related works, the datasets used varies from article to article. Each work splits the dataset randomly into test sets and training sets and sometimes private datasets are being used for evaluation. Therefore, the datasets used represent a comprehensive selection across works. The distribution of images to the training and test splits are summarized in Table 1.

For table detection, all previously mentioned datasets were used as they have ground-truths for tables. Adjustments and consolidation of the labeling done for the private dataset were required. Each dataset's training set was carefully groomed to verify the data encapsulates all *'tabular regions'* instead of just tables with clear boundaries. In the process, most of the Marmot dataset had been eliminated. In the final training set, only 829 images were kept from the Marmot dataset.

For structure detection UNLV was the only dataset which provided annotations for columns and rows, making the data for table structure detection limited. The UNLV dataset consists of scanned documents of varying quality, in some cases tilted images adding difficulty to the dataset. For this project, the focus is not on information extraction from scanned documents, therefore, for accurate table segmentation, several document images were removed from this set due to them containing large amounts of tilt from poor page scans. The task is table detection from documents, which becomes obscured when tilt correction must be created as well. These simplifying assumptions helped focus the project and enabled substantial improvements to the structure detection networks due to noise reduction in the source data. For testing, 50 images are kept aside, and the rest of the dataset is used for training.

The deep neural networks were trained on a combined training dataset created from the public and private datasets. The hand-labeled table dataset consists of more than 2300 document pages from the private dataset, combined with the publicly available data. In total, the training set consists of 5348 training points and the validation set is composed of 171 images. The complete ICDAR 2013 dataset is used, a partial random subset of ICDAR 2017 test set that contains 200 images and randomly selected 50 images from the UNLV dataset to test the proposed methods.

With this combination, the dataset represents a wide range of documents from different fields to test on; ranging between academic writing, newspapers, electronic hardware suppliers, foreign language government documents, and more. This was an important objective since the aim is to solve the information discovery task on the electronic component datasheets and as it can be seen in Fig. 2, documents in the private dataset have unique cases that cannot be found in academic papers or the other specific domains encountered in publicly available datasets. All of these test images are combined and prepared as a 'combined' test set to test the model on every type of document present. The combined test set contains 688 images.

4. Performance study

In this section the performances of each network mentioned prior in Section 3 on the competition datasets of ICDAR 2013 and 2017, UNLV, as well as the private datasets aggregated by Lytica inc. from the e-component specifications sheets available from many manufacturers are explored. All detections were performed using up to four Nvidia P100 GPUs, on a compute node with 32 GB of RAM and 32 CPU cores. Mask-RCNN model requires 8 GB of VRAM which is available on commercial GPUs as well. It is also possible to down-scale the input size if the hardware is not available.

4.1. Training details and experiments

Three distinct deep learning models are trained; an E2E structure detector, table detector, and a standalone structure detector. E2E detector takes a full document image and makes predictions on both table locations as well as the internal table structure. This model does

[1] http://www.icst.pku.edu.cn/cpdp/ICDAR2017_PODCompetition/dataset.html.

[2] http://www.icst.pku.edu.cn/cpdp/data/marmot_data.htm.

not share knowledge between the table and internal structure detector since the detections are done at the same time. Table detector takes an image and outputs predicted table locations. After this step, some post-processing techniques that are explained in Section 3 are applied. The tables are adversarially used against each other and through group consensus, if tables have significant overlap, a combined bounding box is output. The output of the structure detector and the third judge model are also employed to approve or decline the table detection. The structure detector takes an image and outputs detected table structure. Preferably, the model starts with a crop of a table location and that is how it is trained. Hence, the table detector's performance has a strong effect on the structure detector.

Based on the findings that stretching the tables vertically improves the detections (Schreiber et al., 2017; Kara et al., 2019) this pre-processing step is applied. This study employs the same idea but instead of empirically choosing a scaling parameter, it attempts to try and find a mid-way value. A threshold is set for the smallest edge of the bounding box to be 50 pixels long. After processing all the dataset, it was found that to satisfy this threshold, the table dataset should be scaled by 7% vertically and the structure dataset should be scaled by 210% vertically and 4% horizontally on average. This also confirms the previous hypothesis that rows are harder to detect due to them being narrow. To make rows visually salient, images have to be scaled to 2.1x of their original height. Columns usually cover enough of the image and hence, do not require pre-processing most of the time.

Two Mask-RCNN models with ResNet-50 backbone feature extractor, based on Faster-RCNN architecture either with or without pre-training, are tested to observe the effect of pre-training. Pre-training is done on MS COCO dataset. And when the pre-training is not used, the weights are randomly initialized. This paper discusses that pre-training is important since the dataset is small and pre-training allows the network to have a general understanding of images beforehand.

The Mask-RCNN model under study includes FPN, GN, and FCN in addition to underlying Faster-RCNN. Segmentation and detection are performed at the same time as explained in He et al. (2017) to prevent objects from competing. Only one segment is predicted which means an object is detected rather than a class. Model's input size is 1200×2000 with a batch size of 1. The ResNet-101 backbone is not used, as ResNet-50 is capable of extracting the required features and increasing the input image size is more important.

The E2E model is only trained on some part of the UNLV dataset for 20 epochs. Table detection models are only trained for 7 epochs and both pre-training and randomly initialized weights seem to have converged to their best in the same epoch. These two networks are trained on the stretched dataset. The pre-trained structure detection model is first trained on the non-stretched version of the training set for 30 epochs. And after that, training is continued on the stretched dataset for 2 more epoch. In the experiments, this increased the precision in comparison with when it was only trained on the stretched dataset. The structure model input size is 800×1333. Increasing the input size only worsened the performance due to restricted structure dataset size and the increase in the information in the images. The structure detection model without pre-training trained on the first version of the dataset for 20 epochs and then trained for 2 more epochs. Although it reached its best validation accuracy earlier, the lack of pre-training could lead to faster overfitting.

A fixed number of epochs (12 epochs for table detection and 32 for structure detection) are used to stop the training, and at every thousandth step (every step is a mini-batch) validation loss is recorded. The weights which have the lowest validation loss is selected as the final weights. An example figure showing training loss is provided for table detection models in Fig. 10. As it can bee seen in the figure, pre-training allows the loss to go deeper (about 0.01 lower loss in average) and therefore, this work proceeds with using them in all comparisons. The figure also shows the scheduling of the learning rate. The training starts with a learning set of 0.01, and divided to 4 GPUs as advised by He et al. (2017). At steps 26 000 and 40 000, the learning rate is divided by 10. This work conducts a basic search for simple hyper-parameters such as learning rate and FPN parameters. However, future work will include further analysis including other parameters.

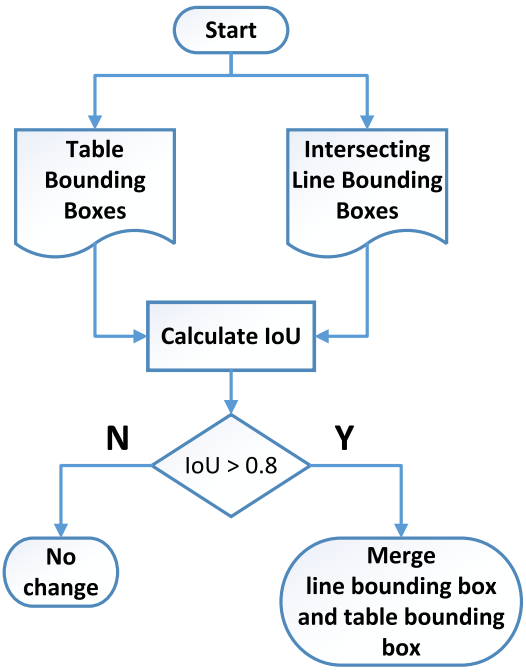


Fig. 9. Agreement module — line agreement algorithm flow.

Different experiments are run to measure the performance of the proposed models. The simplest one is to test the pure E2E model on UNLV dataset. This dataset has full document pages and the expected detections are tables, rows, and columns at the same time. Secondly, the proposed table detection method is tested and compared with other well-known current detectors on ICDAR 2013, ICDAR 2017, UNLV and the private dataset. The proposed post-processing methods are tested with the table detection model. The next experiment includes the generation of table crops from table detector and feeding them into the structure detector to obtain the E2E detection results. The detections are considered correct if the detection has an IoU higher than 0.5 with the ground-truth. Mask-RCNN implementation is based on the Detectron/Caffe2 (Girshick et al., 2018) library using Python.

4.2. Performance comparison

In Table 2, table detection results are presented and compared and in Table 4 structure detection and E2E systems' test results are presented on public datasets. In Table 3, the same models are tested on the private dataset and the combined test set. Mean average recall, precision and F-measurements for this configuration. The IoU threshold for a detection to be considered true is set to 0.5 to be able to compare the detection results with prior research. In order to present a comparison, before going into detection details, it is beneficial to discuss execution lengths of the models. The proposed Mask-RCNN based deep learning model with ResNet-50 backbone network takes 0.6 s (on average) to infer a single image on a single P100 GPU and up to 2 s on certain images (where the images are stretched in structure detection). However, in Traquair et al. (2019), authors achieve processing time of 0.2 s with RetinaNet and 0.4 s with Faster-RCNN with ResNet-50 backbone network. While Mask-RCNN is comparably slower than the Faster-RCNN and RetinaNet, it outperforms these models. Here it is a good spot to talk about one other downside of the stitched E2E

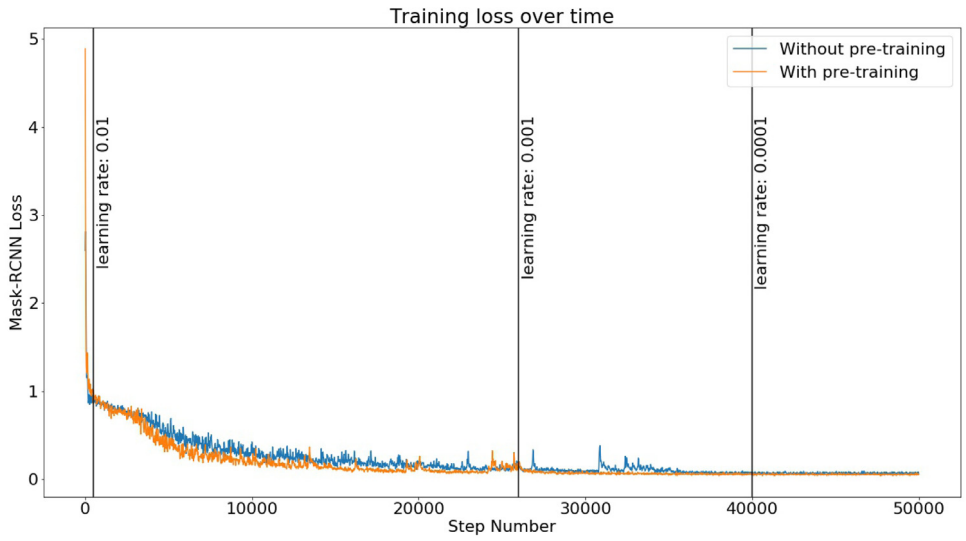


Fig. 10. Training loss over time for the Mask-RCNN models.

Table 2
Table detection performance on public datasets.

Methods	ICDAR 2013			ICDAR 2017		
	Recall	Precision	F1-Score	Recall	Precision	F1-Score
Proposed (A) – Without pre-training	**1.0000**	0.9662	0.9828	**1.0000**	0.9721	0.9858
Proposed (A) – With pre-training	**1.0000**	0.9900	0.9950	**1.0000**	0.9868	0.9933
Proposed (A) + Judge & Agreement	**1.0000**	**0.9905**	**0.9952**	**1.0000**	**0.9876**	**0.9937**
Traquair et al. (2019)	0.9808	0.9738	0.9773	0.9685	0.9385	0.9533
Kavasidis et al. (2018)	0.9810	0.9750	0.9780	–	–	–
Schreiber et al. (2017)	0.9615	0.9740	0.9677	–	–	–
Tran et al. (2015)	0.9636	0.9521	0.9578	–	–	–
Hao et al. (2016)	0.9215	0.9724	0.9463	–	–	–

UNLV has been used for both table detection and internal structure detection. Model A represents the pre-trained Mask-RCNN based table detection model. Judge and Agreement are the proposed improvement approaches. The primary aim of the improvements was to eliminate false-positives and increase precision. The missing values are due to the related papers not have performed evaluations on those datasets.

Table 3
Table detection on public UNLV dataset and the private dataset.

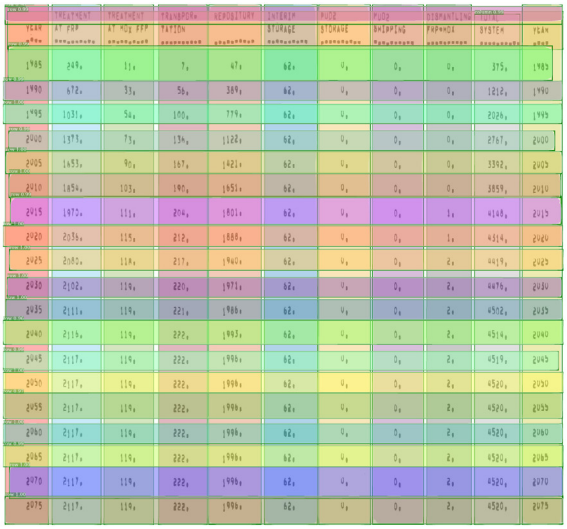
Methods	Private			UNLV		
	Recall	Precision	F1-Score	Recall	Precision	F1-Score
Proposed (A) – Without pre-training	**1.0000**	0.9854	0.9926	**1.0000**	0.9797	0.9897
Proposed (A) – With pre-training	1.0000	0.9980	0.9990	**1.0000**	0.9819	0.9909
Proposed (A) + Judge & Agreement	**1.0000**	**0.9980**	**0.9990**	**1.0000**	**0.9820**	**0.9910**
Gilani et al. (2017)	–	–	–	0.9067	0.8230	0.8629
Shafait and Smith (2010)	–	–	–	0.7900	0.8600	0.8200

networks. Since they use multiple passes from the Mask-RCNN models, they require up to 5 s for a single image on a single GPU.

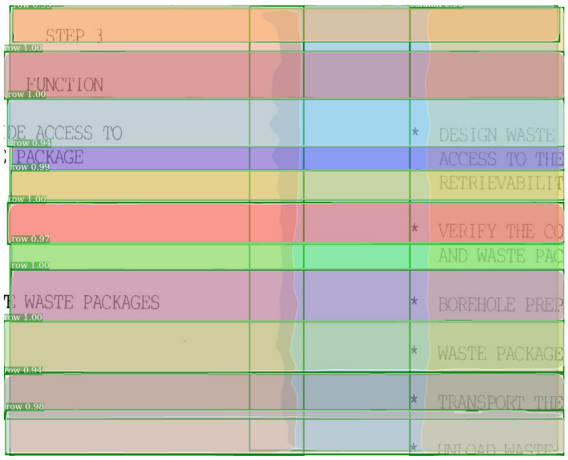
The proposed table detection model (Model A) is achieving the new state-of-the-art recall and precision on multiple datasets. It is also evident that although the model without pre-trained weights achieves a high success, it is not as successful as the pre-trained weights. The same phenomenon is also observed in table structure detection model (Model B). Pre-training is one of the key factors to have more robust object detectors. However, even without pre-training, the Mask-RCNN method proves to be successful and with increased amounts of data, the pre-training stage can be eliminated.

Previous research shows their result mostly only on ICDAR 2013 however other datasets such as UNLV are more challenging and the results show the performance of the proposed models on each accordingly. Additionally, although this paper compares results with others, training is performed on randomly sampled instances drawn from an aggregated pooled set of multiple datasets and the results are not exactly produced by replicating the datasets employed in previous research (Schreiber et al., 2017; Kavasidis et al., 2018) because they train on random splits of either public or private datasets. Therefore, the training and testing is performed on a random combination of public and private datasets. This work aims to show how deep learning can be used in supply chain optimization and provide an insight on how effective it is. And to do this, the primary objective is to show how effective Mask-RCNN can be on this problem.

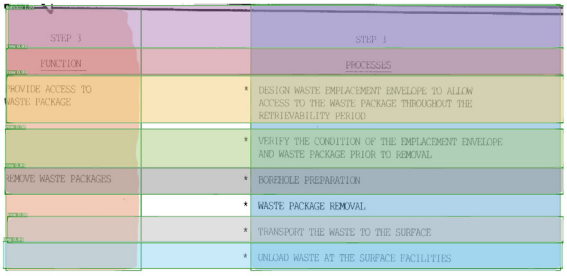
The proposed table detector is tested with and without the judge and agreement enhancement modules. The addition of post-processing eliminates false-positives and increases the precision in most cases but sometimes the metrics are not affected as one would have expected, such as it was in ICDAR 2017. This is because the judge takes as an input; (1) the table detections (which is quite accurate) and (2) the outputs from E2E structure detection model (which is not as accurate). Therefore, in some difficult cases, the judge module eliminates true-positives as well. After the visual inspection of results, it is evident that most of the false-positives are caused by figures that look like tables, as one example can be seen in Fig. 5a. Labeling and including figures



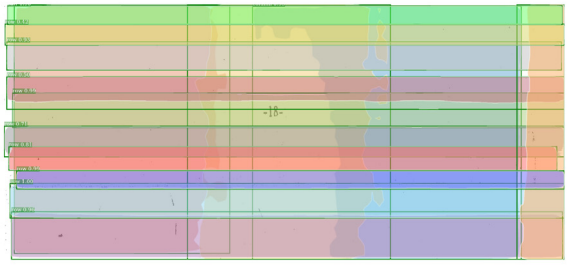
(a) Accurate table and structure detection



(b) Incomplete table but structure is still detected



(c) Complete version of the table in Fig. 11b



(d) False-positive table detection. But structure is detected

Fig. 11. E2E detection examples for proposed models.

and formulas to the dataset along with tables might help increase the performance.

Structure detection model (Model B) accepts table crops as input and detects structure detections. Therefore, its performance is purely dependent on the table detection model's performance. As it can be seen in the E2E segment of Table 4, imprecise detections by the model A lowers the performance for the E2E system. An example is given in Fig. 5b where an incomplete table detection causes false-positives in structure detection. It should be acknowledged that there are many solutions to perfect this system, such as adding a second layer to the Model A to confirm the detections via a binary table classifier. The E2E structure detection model (Model C) takes the document page images as input and provides as output the locations of rows and columns. This is performed in a single pass but the results are not up to par. In particular, row detection results have suffered, even though all the images are stretched. E2E table and structure detection model (Model D) outputs tables and structures in a single pass. This model increases performance and surpassed all the other E2E methods. Addition of the tables to the Model C helped the deep learning model to learn where tables are. Hence, higher precision and recall on rows and columns are observed.

The E2E networks (employing model A + B), or Vanilla Two-Stage System, starts with an image, then outputs a table with or without the improvements and then this table crop is fed into the Model B which outputs internal structure locations. Model B's performance is reduced by imperfect detections or false-positives output from Model A, which drop precision notably. Because Model B takes table crops as input and outputs structures, during the training process, the model learned that every crop consists of many rows and columns that start at the left side of the image and ends at the right side of the image (or top and bottom for columns). Therefore, if there is a false-positive in table detection,

Table 4
Structure detection and E2E performance of the proposed systems.

Methods	UNLV		
	Recall	Precision	F1-Score
Structure detection			
Proposed (B) — Without pre-training	0.9536	0.9420	0.0.9478
Proposed (B) — With pre-training	**0.9853**	**0.9759**	**0.9606**
E2E detection			
Proposed (C)	0.8675	0.7913	0.8276
Proposed (D)	0.9224	**0.8660**	**0.8942**
Proposed (A+B)	0.9615	0.7001	0.8308
Proposed (A + B) + Judge & Agree.	**0.9615**	0.7140	0.8378

this model again predicted rows and columns (usually text lines are detected as rows) and therefore generating a large number of false-positives which reduces the precision by a significant amount. This phenomenon can be observed in Fig. 11d. Additionally, as illustrated in Fig. 4b, since there are cases where tables are contained within tables, it is not easy to eliminate some of the false-positives because no table should be missed. Instead, some drop in precision is preferred. The proposed model is able to detect cases where a table inside another table (inside a row) as can be seen in Fig. 4b. However, further adjustments over the parameters in 3.3 causes the model to eliminate the smaller table.

Table 3 shows that a high performance on the private dataset and combined dataset is achieved by the proposed models. Achieving a recall of 1.0 in all of the datasets with a precision of about 0.99 in table detection. Also, as Fig. 2 demonstrates, the private dataset presents a higher difficulty compared to ICDAR and UNLV datasets. These public datasets are collected mostly from academic documents

and government documents. Therefore, they are strictly regulated and stylistically easy to follow. Whereas the dataset under study is a collection of datasheets from diverse manufacturers and manifests unique characteristics. In addition to these, the combined test set is used to be able to benchmark models on a wide-range of tables from different domains to simulate real-world use cases.

5. Conclusion and future directions

A vital element in supply chains is emerging as business processes are moving to digital mediums. Information flow from supplier to consumers is larger than ever, requiring faster and more efficient processing of documents containing vital information. As a prominent and dense container of information, tabular data is vital to extract as it contains data which standard optical character recognition (OCR) utilities fail to identify and relate in any meaningful way. As prior work has gone into the detection of tables (Traquair et al., 2019; Kavasidis et al., 2018; Fang et al., 2011; Shafait and Smith, 2010; Gilani et al., 2017) as well as into the structure detection (Kara et al., 2019; Schreiber et al., 2017), this paper fills in the gaps and proposed three systems by which these methods could be used in unison to optimize for maximum content coverage.

Having two consensus-based algorithms, the judge system and the agreement system aims to prioritize table detection precision to address the propagation of errors from poor table detections. The vanilla system of raw table detections to structure detection relies entirely on the accuracy of the CNNs with no opportunity for adjustments. Each model performed near identically on the available datasets of ICDAR, with fractions of decimals indifference. Though each model is so individually accurate, the E2E methods each suffered from the number of false-positives. With already achieved near-perfect precision for tables and with table structure data labeling, the table structure network may perform equally. Future work can go into filtering table detections, or training structure detection on splits of no structure images, such that it may itself filter the table detections. In addition to these,broadening the list of objects detected, such as including figures, may also allow for greater segmentation of discrete components of documents, further increasing the E2E accuracy by refining what is classified as a table.

CRediT authorship contribution statement

Ertugrul Kara: Validation, Methodology, Writing - original draft. **Mark Traquair:** Validation, Methodology, Writing - original draft. **Murat Simsek:** Methodology, Writing - review & editing. **Burak Kantarci:** Investigation,Methodology, Supervision, Writing - review & editing, Funding acquisition. **Shahzad Khan:** Methodology, Supervision, Writing - review & editing.

References

Canny, J., 1987. A computational approach to edge detection. In: Readings in Computer Vision. Elsevier, pp. 184–203.

Cesarini, F., Marinai, S., Sarti, L., Soda, G., 2002. Trainable table location in document images. In: Object Recognition Supported by User Interaction for Service Robots, Vol. 3. IEEE, pp. 236–240.

Fang, J., Gao, L., Bai, K., Qiu, R., Tao, X., Tang, Z., 2011. A table detection method for multipage pdf documents via visual seperators and tabular structures. In: 2011 International Conference on Document Analysis and Recognition. IEEE, pp. 779–783.

Gilani, A., Qasim, S.R., Malik, I., Shafait, F., 2017. Table detection using deep learning.

Girshick, R., 2015. Fast r-cnn. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1440–1448.

Girshick, R., Radosavovic, I., Gkioxari, G., Dollár, P., He, K., 2018. Detectron. https://github.com/facebookresearch/detectron.

Hao, L., Gao, L., Yi, X., Tang, Z., 2016. A table detection method for pdf documents based on convolutional neural networks. In: 2016 12th IAPR Workshop on Document Analysis Systems (DAS). IEEE, pp. 287–292.

He, K., Gkioxari, G., Dollar, P., Girshick, R., 2017. Mask r-cnn. In: 2017 IEEE International Conference on Computer Vision (ICCV). URL http://dx.doi.org/10.1109/ICCV.2017.322.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778.

Hu, P., Ramanan, D., 2017. Finding tiny faces. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 951–959.

Ioffe, S., Szegedy, C., 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167.

Kara, E., Traquair, M., Kantarci, B., Khan, S., 2019. Deep learning for recognizing the anatomy of tables on datasheets. In: IEEE Symposium on Computers and Communications (ISCC). Barcelona, Spain, (Accepted) Available for reviewer: http://nextconlab.academy/iscc19/1570532191.pdf.

Karatzas, D., Shafait, F., Uchida, S., Iwamura, M., i Bigorda, L.G., Mestre, S.R., Mas, J., Mota, D.F., Almazan, J.A., De Las Heras, L.P., 2013. Icdar 2013 robust reading competition. In: Intl. Conf. on Document Analysis and Recognition (ICDAR). IEEE, pp. 1484–1493.

Kavasidis, I., Palazzo, S., Spampinato, C., Pino, C., Giordano, D., Giuffrida, D., Messina, P., 2018. A saliency-based convolutional neural network for table and chart detection in digitized documents. arXiv preprint arXiv:1804.06236.

Kieninger, T., Dengel, A., 1998. The t-recs table recognition and analysis system. In: International Workshop on Document Analysis Systems. Springer, pp. 255–270.

Kiryati, N., Eldar, Y., Bruckstein, A.M., 1991. A probabilistic hough transform. Pattern Recognit. 24 (4), 303–316.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems. pp. 1097–1105.

Lin, T.-Y., Dollár, P., Girshick, R.B., He, K., Hariharan, B., Belongie, S.J., 2017a. Feature pyramid networks for object detection. In: CVPR, Vol. 1. p. 4.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollar, P., 2017b. Focal loss for dense object detection. In: 2017 IEEE International Conference on Computer Vision (ICCV). URL http://dx.doi.org/10.1109/ICCV.2017.324.

Mao, S., Rosenfeld, A., Kanungo, T., 2003. Document structure analysis algorithms: a literature survey. In: Document Recognition and Retrieval X, Vol. 5010. International Society for Optics and Photonics, pp. 197–208.

Milosevic, N., Gregson, C., Hernandez, R., Nenadic, G., 2019. A framework for information extraction from tables in biomedical literature. Int. J. Doc. Anal. Recognit. (IJDAR) 22 (1), 55–78.

Oro, E., Ruffolo, M., 2009. Trex: An approach for recognizing and extracting tables from pdf documents. In: 2009 10th International Conference on Document Analysis and Recognition. IEEE, pp. 906–910.

Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. In: Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1. NIPS'15. MIT Press, Cambridge, MA, USA, pp. 91–99.

Rice, S.V., Jenkins, F.R., Nartker, T., 2012. The fourth annual test of ocr accuracy.

Schreiber, S., Agne, S., Wolf, I., Dengel, A., Ahmed, S., 2017. Deepdesrt: Deep learning for detection and structure recognition of tables in document images. In: 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), Vol. 01. pp. 1162–1167.

Serra, J., 1983. Image Analysis and Mathematical Morphology. Academic Press, Inc., Orlando, FL, USA.

Shafait, F., Smith, R., 2010. Table detection in heterogeneous documents. In: Proceedings of the 9th IAPR International Workshop on Document Analysis Systems. ACM, pp. 65–72.

Shahab, A., Shafait, F., Kieninger, T., Dengel, A., 2010. An open approach towards the benchmarking of table structure recognition systems. In: ACM International Conference Proceeding Series. pp. 113–120.

Shelhamer, E., Long, J., Darrell, T., 2017. Fully convolutional networks for semantic segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 39 (4), 640–651, URL http://dx.doi.org/10.1109/TPAMI.2016.2572683.

Shin, H.-C., Roth, H.R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., Summers, R.M., 2016. Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. IEEE Trans. Med. Imaging 35 (5), 1285–1298.

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

Tjahjono, B., Esplugues, C., Ares, E., Pelaez, G., 2017. What does industry 4.0 mean to supply chain? Procedia Manuf. 13, 1175–1182, manufacturing Engineering Society International Conference 2017, MESIC 2017, 28–30 June 2017, Vigo (Pontevedra), Spain.

Tkaczyk, D., Szostek, P., Fedoryszak, M., Dendek, P.J., Bolikowski, Ł., 2015. Cermine: automatic extraction of structured metadata from scientific literature. Int. J. Doc. Anal. Recognit. (IJDAR) 18 (4), 317–335.

Tran, D.N., Tran, T.A., Oh, A., Kim, S.H., Na, I.S., 2015. Table detection from document image using vertical arrangement of text blocks. Int. J. Contents 11 (4), 77–85.

Traquair, M., Kara, E., Kantarci, B., Khan, S., 2019. Deep learning for the detection of tabular information from electronic component datasheets. In: IEEE Symposium on Computers and Communications (ISCC). Barcelona, Spain, (Accepted) Available for reviewer: http://nextconlab.academy/iscc19/1570516006.pdf.

Vaidya, S., Ambad, P., Bhosle, S., 2018. Industry 4.0 – a glimpse. Procedia Manuf. 20, 233–238, 2nd International Conference on Materials, Manufacturing and Design Engineering (iCMMD2017), 11–12 December 2017, MIT Aurangabad, Maharashtra, INDIA.

Wu, Y., He, K., 2018. Group normalization. arXiv preprint arXiv:1803.08494.