Novosibirsk State University
*THE REAL SCIENCE

Department of Mechanics & Mathematics

TableNet: Deep Learning model for End-to-End Table detection and Tabular Data Extraction from Scanned Document Images

Novosibirsk State University

Muhammad Hami Asmai Bin Ismail

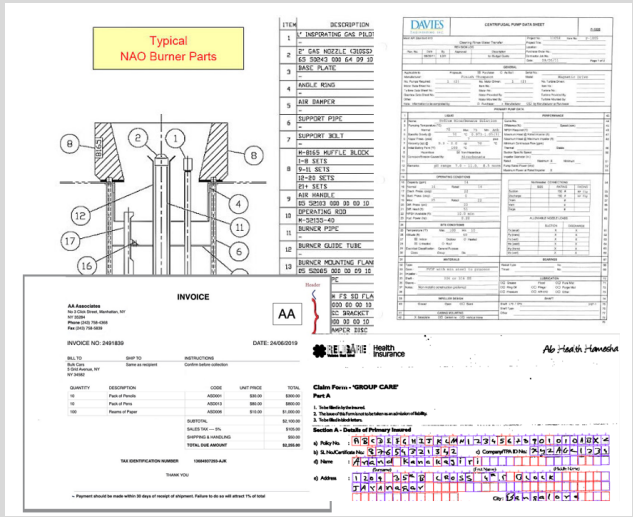
Advisor: Prof. Dmitry Tailakov,. Digital Field Technologies

Objective: The presentation is to discuss in detail, about a novel paper written by Shubham Paliwal, Vishwanath D, Rohit Rahul, Monika Sharma and Lovekesh Vig (TCS Research, New Delhi)

General Problem

The Big Picture

- Unstructured documents.
- Tabular information.
- Examples:
  - Hand-written forms.
  - Invoices.
  - Technical datasheets.
  - Bill of Materials.
  - Others.



Sub-Problems

Tabular data detection can be broken down into 2 sub-problems

1 - Table Detection

Recognize and locate the table.

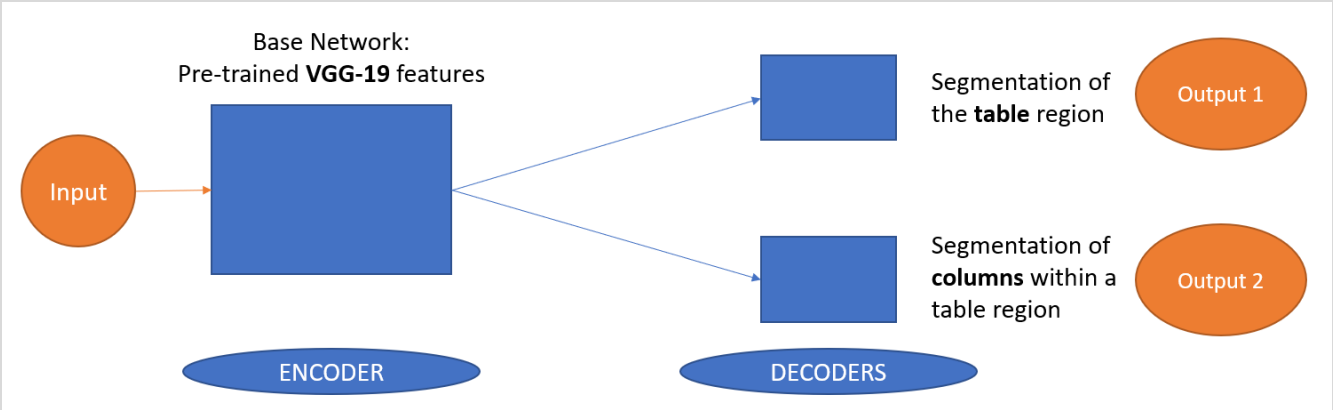
2 - Table Structure Recognition

Identify the columns and data inside.

Model Implementation

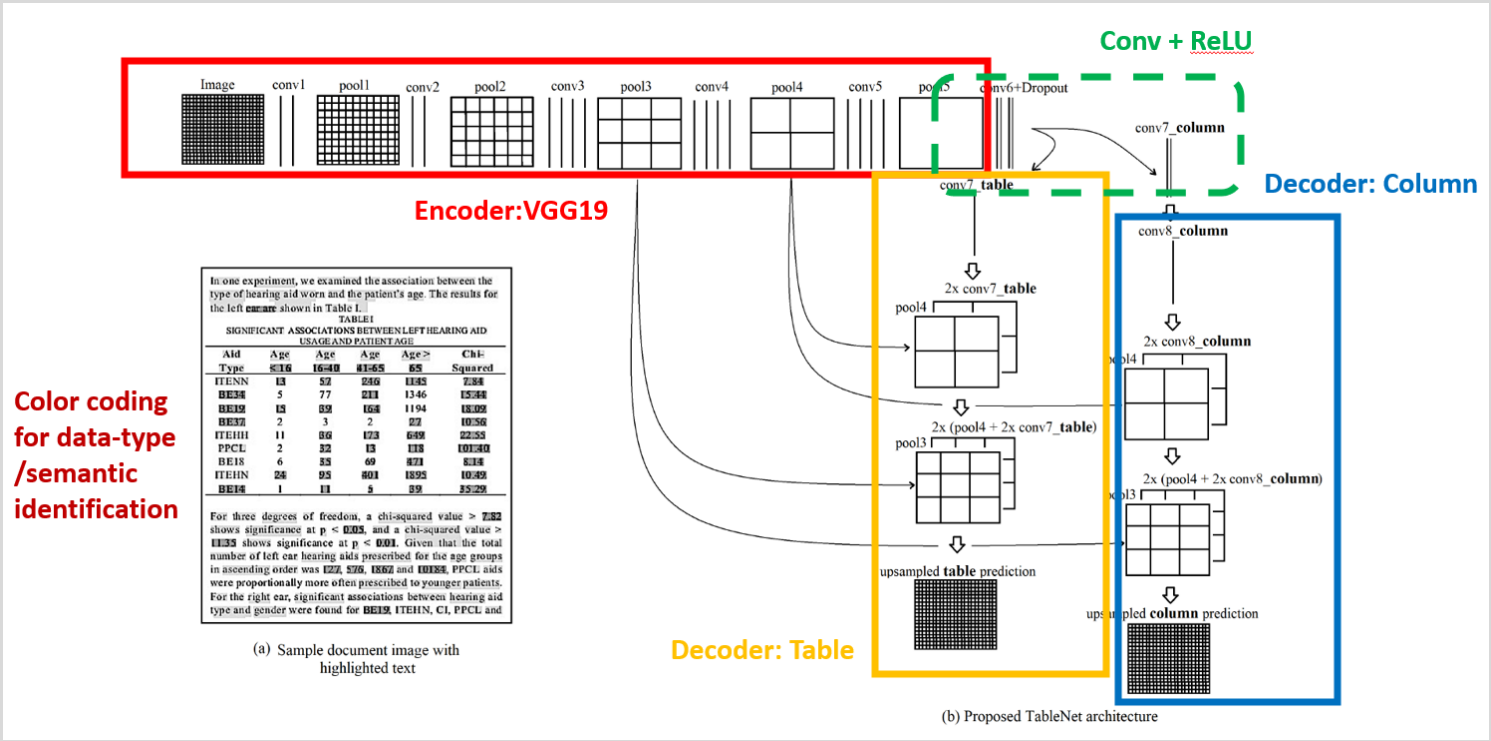
General model representation

- The aim is to solve the 2 sub-problems simultaneously, unlike most other models.
- The implementation is based on encoder-decoder model using multi-task approach.



Detailed Architecture

End-to-end architecture of TableNet



In one experiment, we examined the association between the type of hearing aid worn and the patient's age. The results for the left ear are shown in Table I.

TABLE I
SIGNIFICANT ASSOCIATIONS BETWEEN LEFT HEARING AID USAGE AND PATIENT AGE

Aid Type	Age ≤16	Age 16-40	Age 41-65	Age >65	Chi-Squared
ITENN	13	57	246	1145	7.84
BE34	5	77	211	1346	15.44
BE19	15	89	164	1194	18.09
BE37	2	3	2	22	10.56
ITEHH	11	86	173	649	22.55
PPCL	2	32	13	118	101.40
BE18	6	35	69	471	8.14
ITEHN	24	95	401	1895	10.49
BE14	1	11	5	89	35.29

For three degrees of freedom, a chi-squared value > 7.82 shows significance at $p < 0.05$, and a chi-squared value > 11.35 shows significance at $p < 0.01$. Given that the total number of left ear hearing aids prescribed for the age groups in ascending order was 127, 576, 1862 and 10184, PPCL aids were proportionally more often prescribed to younger patients. For the right ear, significant associations between hearing aid type and gender were found for BE19, ITEHN, CI, PPCL and

(a) Sample document image with highlighted text

(b) Proposed TableNet architecture

Department of Mechanics & Mathematics

Data Extraction

How the data extracted from tables and columns?

Generated Masks

used to filter out the table column regions.

Tesseract OCR

used to extract texts inside the table column regions.

Regular Expressions

used to determine data type or semantic feature.

Row Segmentation

How the rows are identified?

3 Row Segmentation Rules

1. In most tables for which line demarcations are present, the line = row.
2. If a row spans multiple lines, the line with max non-blank columns = new row.
3. If all columns filled and no line demarcations, again each line = row.

	Part Number	Short Description	Long Description
Start of row 1	X12345	PUMP, CENTRIFUGAL, SINGLE STAGE	Unused Surplus SPX Clyde Union FK14/18, API 610, OH2 pump with the following features: • Case material • Single stage • impeller • suction flange • discharge flange • CW Rotation • Weight 8,500 lbs • Dims 104"Lx104"Wx98"H
Start of row 2	342-99A	VALVE,BALL,2"X1500#, FLOATING, LEVER, SS316	1" CLASS 1500 FLOATING BALL VALVE BOLTED BONNET, TWO-PIECE BODY, FLOATING BALL FULL BORE, BLOWOUT PROOF STEM FIRE SAFE AND ANTI STATIC STAINLESS STEEL BALL ASTM A216 WCB, A351 CF8, A351 CF8M, A351 CF3, A351 CF3M, A351 CN7M

Experiment Parameters

Dataset and Computing Power

Dataset

Marmot table recognition dataset - total of 1016 documents (509 English documents), annotated for training.

Computing Power

Intel(R) Xeon(R) Silver CPU having 32 cores and RAM of 128 GB Tesla V100-PCIE-1 GPU with 6GB of GPU memory.

Experiment Parameters

Running Parameters by Phases

- Initial Phase:
  - Table to column computation ratio = 2:1.
  - 500 iterations with batch size of 2.
- Next Phases:
  - Table to column computation ratio = 1:1.
  - 5000 iterations with batch size of 2, learning rate 0.0001.
  - Use of Adam Optimizer (Beta1=0.9, Beta2=0.999, Epsilon=1e-08).

Experiment Result

Comparative table of Score

Model	Recall	Precision	F1-Score
TableNet + Semantic Features (fine-tuned on ICDAR)	**0.9628**	**0.9697**	**0.9662**
TableNet + Semantic Features	**0.9621**	**0.9547**	**0.9583**
TableNet	**0.9501**	**0.9547**	**0.9547**
DeepDeSRT [8]	0.9615	0.9740	0.9677
Tran et al [10]	0.9636	0.9521	0.9578

TABLE I: Results on Table Detection

Model	Recall	Precision	F1-Score
TableNet + Semantic Features (fine-tuned on ICDAR)	**0.9001**	**0.9307**	**0.9151**
TableNet + Semantic Features	**0.8994**	**0.9255**	**0.9122**
TableNet	**0.8987**	**0.9215**	**0.9098**
DeepDeSRT [8]	0.8736	0.9593	0.9144

TABLE II: Results on Table Structure Recognition & Data Extraction

Conclusion

Conclusion of the experiment

- TableNet is a novel deep learning model trained on dual tasks of table detection and structure recognition in an end-to-end fashion.
- TableNet is the first model that solves the two sub-problems simultaneously.
- Data type or semantic feature identification can improve accuracy.

Future Works

Possible improvements identified by the author

- Possible expansion of the algorithm by introduction of a third branch to include row identification.
  - This requires manual annotation exercise.
- Inclusion of other type of semantic knowledge.
  - Example: currency, country, city.