Table Extraction

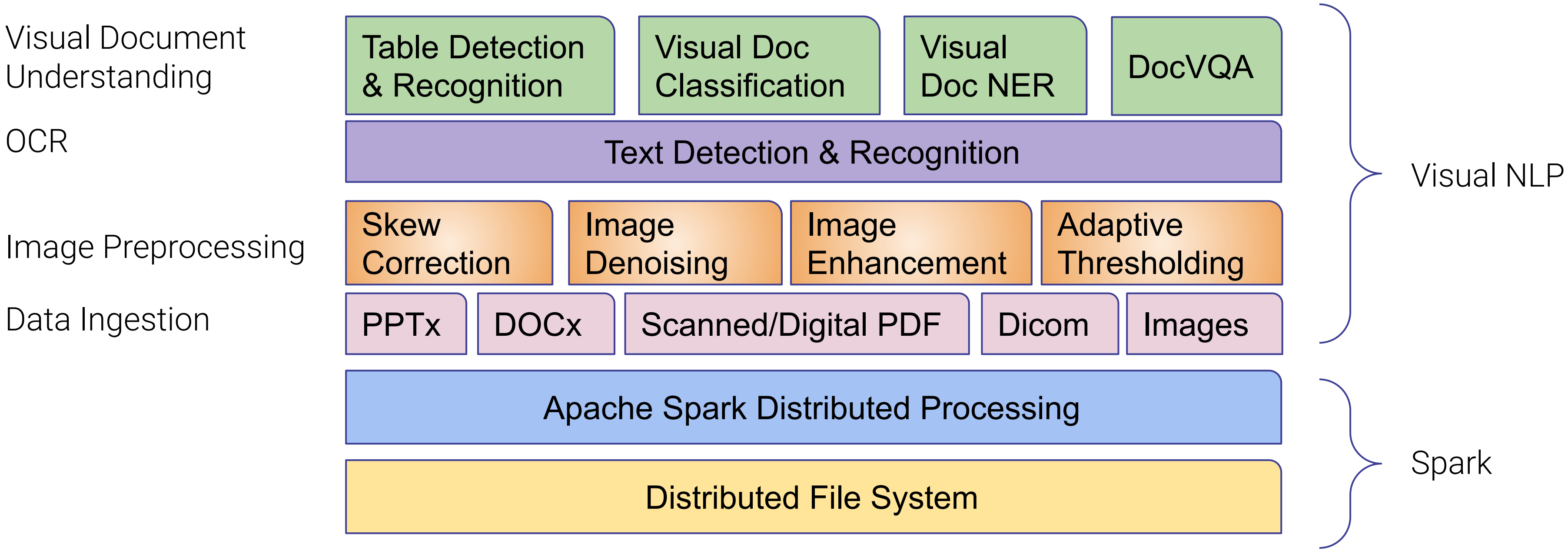
for Data Scientists

Visual NLP Team, John Snow Labs

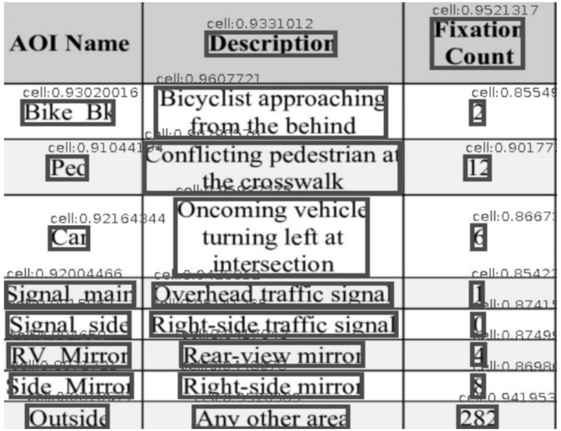
Agenda

Main topic	Introduced Concepts
Introduction	Visual-NLP brief intro. What is Table Processing. Examples.
General Pipeline Architecture	Different architectures; table & cell detection, OCR. Examples.
Alternative Approach	Visual Question Answering. Problem definition. Extracting information from Tables. Examples.
Summary and next steps	JSL's roadmap on Table Extraction.
Questions & Answers	Questions to discuss the content.

Visual NLP


Visual Document Understanding
Table Detection & Recognition
Visual Doc Classification
Visual Doc NER
DocVQA
OCR
Text Detection & Recognition
Image Preprocessing
Skew Correction
Image Denoising
Image Enhancement
Adaptive Thresholding
Data Ingestion
PPTx
DOCx
Scanned/Digital PDF
Dicom
Images
Apache Spark Distributed Processing
Distributed File System
Visual NLP
Spark

Introduction

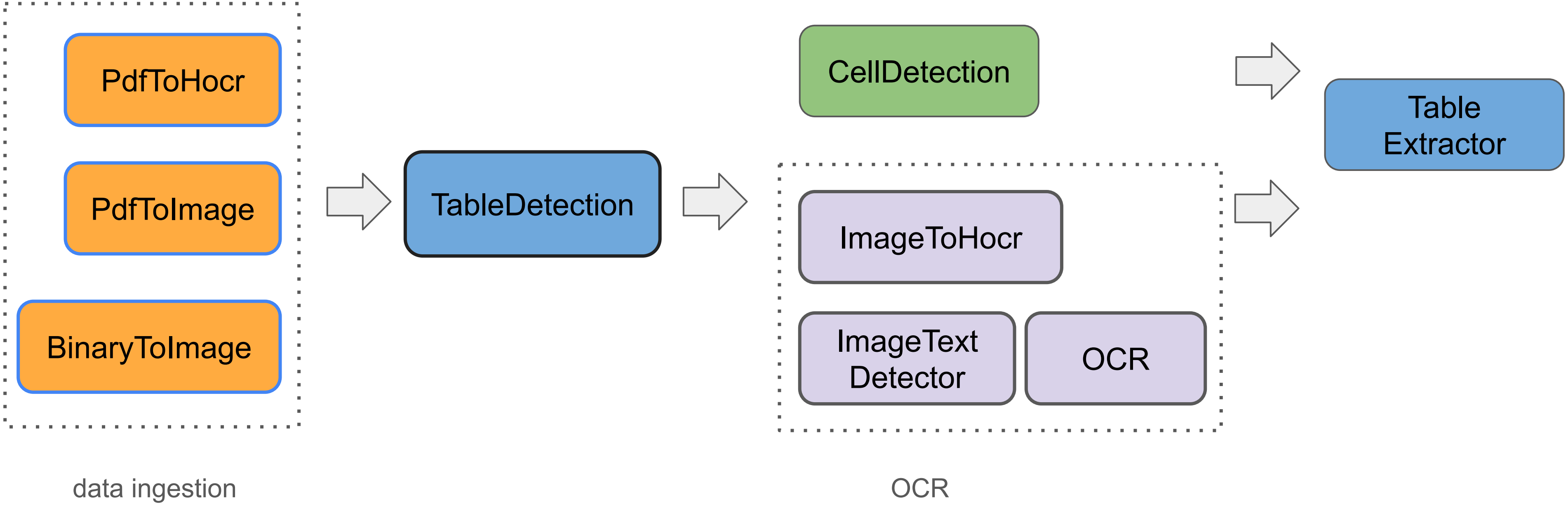


AOI Name	Description	Fixation Count
Bike_Bk	Bicyclist approaching from the behind	2
Ped	Conflicting pedestrian at the crosswalk	12
Car	Oncoming vehicle turning left at intersection	6
Signal_main	Overhead traffic signal	1
Signal_side	Right-side traffic signal	0
RV_Mirror	Rear-view mirror	4
Side_Mirror	Right-side mirror	8
Outside	Any other area	282

col0	col1	col2
Empty	Description	Fixation cm
Bike_Bk	Bieyelis! approaching	2
Ped	(Conflicting pedestrian at| the crosswalk	RD
Car	Oncoming vehicle turning left at intersection	6
Signal_main|	Overhead traffic signal	1
Signal_side	Right-side traffic signal	0
RV_Mirror	Rear-view mirror Right-side mirror	4 8
Outside	Any other area	282

- Table Extraction -> recover the structure, hierarchies and relations of elements in a table.
- Involves multiple steps, OCR, table detection, cell detection.

Pipeline Architecture



Metrics

1.1. Table Detection Benchmark (Icdar 2019)

Evaluation on icdar19 cTDaR Track A dataset:

Model	**0.6** (IOU F1)	**0.7** (IOU F1)	**0.8** (IOU F1)	**0.9** (IOU F1)	**Weighted Average**
icdar_13_td	0.839	0.820	0.815	0.769	0.807
icdar_19_track_a_modern_td	0.918	0.898	0.893	0.837	0.882
table_bank_both_td	0.839	0.820	0.815	0.769	0.807
icdar_19_track_a_modern_td_v2	0.918	0.898	0.893	0.837	0.882
general_model_td_v2	0.918	0.900	0.893	0.839	0.883
table_detection_v3	**0.951**	**0.946**	**0.936**	0.887	**0.93**
CascadeTabNet (general model)	0.943	0.934	0.925	**0.901**	0.923
DETR_td	0.580	0.560	0.510	0.306	0.474

* All models have been fine tunned in CTDAR Track A except DETR_td and DETR_td (FT in PubTables-1M)

2. Table Detection Benchmark (Icdar 13)

Evaluation on icdar13 dataset:

Model	**0.6** (IOU F1)	**0.7** (IOU F1)	**0.8** (IOU F1)
DeepDeSRT	0.9615	0.974	0.9677
TableNet	0.9628	0.9697	0.9662
general_model_td_v2	**1.0**	**1.0**	**1.0**

* All models have been fine tunned in CTDAR Track A except DETR_td and DETR_td_v2 (FT in PubTables-1M)

2. Table Structure Recognition (TSR)

2.1. TSR in ICDAR 19 TRACK B

Model	**0.6** (IOU F1)	**0.7** (IOU F1)	**0.8** (IOU F1)	**0.9** (IOU F1)	**Weighted Average**
CascadeTabNet	0.438	0.354	0.19	0.036	0.232
NLPR-PAL	0.365	0.305	0.195	0.035	0.206
Multi-Type-TD-TSR	**0.589**	**0.404**	0.137	0.015	**0.253**
SparkOCR	0.306	0.289	**0.253**	**0.097**	0.225

* DETR not included because is trained in PubTables-1M. TSR accuracy in Track B low

2.2. TSR in PubTables-1M

Test performance of TSR models on PubTables-1M (Paper).

Model	AP	AP50	AP75	AR
Faster R-CNN	0.722	0.815	0.785	0.762
DETR	**0.912**	**0.971**	**0.948**	**0.942**

* Performance on different subsets (AP, AP50, AP75 and AR) of PubTables-1M dataset

3. Table Cell Detection

TSR in ICDAR 19 TRACK B with TedEval results

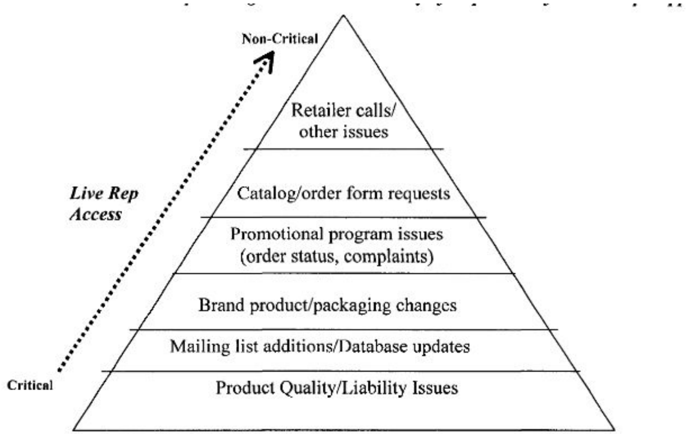
Model	Recall	Precision	Hmean
CascadeTabNet			
region_cell_detection	0.925	0.846	0.884

hon script.py -g=../.Datasets/Table\ Cell\ Detection/trackb2_adapted_gt.zip -s=../.Datasets/Table\ Cell\ Detection/tsr_cd_results_trackb2_adapted.zip
ulated!{"recall": 0.925943270002676, "precision": 0.8463210088148874, "hmean": 0.8843435504967244, "AP": 0}(base)

Alternative DocVQA

- Many times extracting the entire table is not crucial.
- DocVQA can be used to process tables and extract specific information.

DocVQA



- *Examined 4 levels of service options ranging from $1.1MM to $6.1MM.*

What is the issue at the top of the pyramid? Retailer calls/ other issues

Which is the least critical issue for live rep support? Retailer calls/other issues

Which is the most critical issue for live rep support? Product quality/liability issues

Not only extract and interpret the textual (handwritten, typewritten or printed) content of the document images, but also other visual cues including layout (page structure, forms, tables), non-textual elements (marks, tick boxes, separators, diagrams) and style (font, colours, highlighting).

Example #1

126

INTERPRETATIVE GUIDE FOR BIOCHEMICAL DATA

GUIDE USED IN INTERPRETATION OF BLOOD DATA - YOUNG ADULT MALES

	Deficient	Low	Acceptable	High
Hemoglobin gm/100 ml	< 12.0	12.0-13.9	14.0-14.9	≥ 15.0
Total Serum Protein gm/100 ml	< 6.0	6.0-6.39	6.4-7.1 1/	≥ 7.2 1/
Serum Vitamin A mcg/100 ml	< 10	10-19	20-49	≥ 50
Serum Carotene mcg/100 ml	< 20	20-39	40-99	≥ 100
Serum Ascorbic Acid mg/100 ml	< 0.10	.10-.19	0.20-0.39	≥ .40

GUIDE USED IN INTERPRETATION OF URINE DATA

What is the Acceptable Haemoglobin level(g/100ml)?

14.0-14.9

What is the deficiency level for Haemoglobin in blood?

<12.0

What is the acceptable level of Serum Carotene in blood?

40-99

Example #2

CODE 617C 7/21/82

SCHOOL LUNCH COOKED SAUSAGE PIZZA

COMPONENT	WEIGHT
Shell 3.2" x 5" (thin formula1)	1.40 oz.
Sauce 101	0.98 oz.
Meat 225	0.60 oz.
Cheese 564	1.52 oz.
NET WEIGHT	4.50 oz.

What is the number circled?

0.98 oz.

What is the net weight?

Retailer calls/other issues

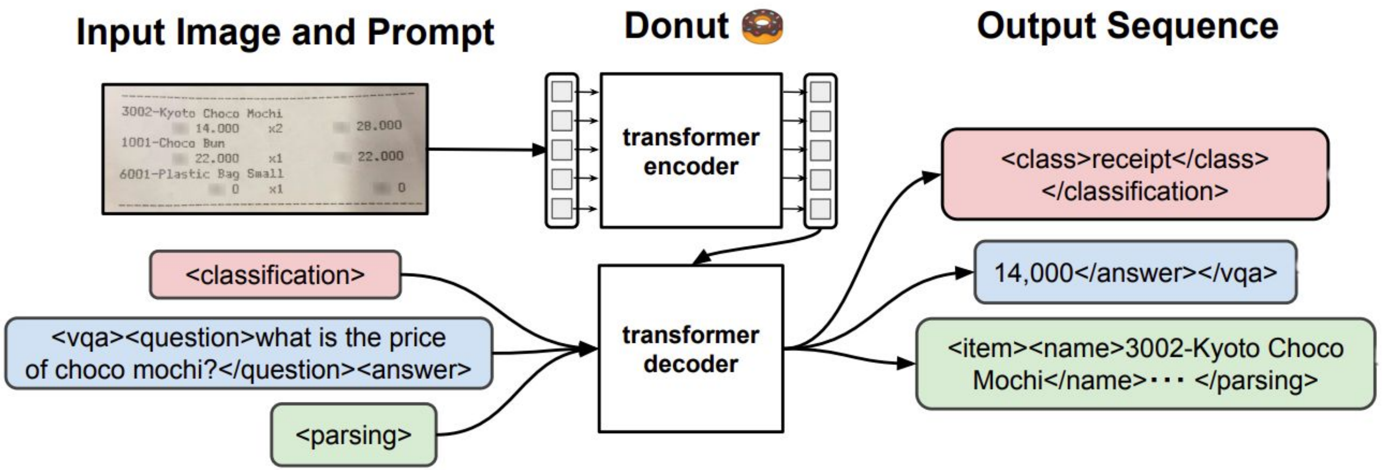
What is the title of the table?

SCHOOL LAUNCH COOKED SAUSAGE PIZZA

Common Architectures

1. **Roughly 2 types of architectures:** OCR + decoder and encoder/decoder.
2. **OCR + decoder:** They use 3 types of features: layout, text, and image.
3. **Encoder-decoder:** they use a visual transformer + language model decoding.

Donut 🍩



- OCR-free VDU model
- Swin Transformer is used for the encoder.
- BART is used as the language decoder.

Best Practices

- Identify the documents/sections you care about.
- Apply a limited set of well defined questions covering the information you need.
- Rephrase the questions.
- Use confidence scores.

Examples

[TableExtractionBasics]

[TableProcessingVQA]

Summary and next steps

We’ve covered…

- Table Extraction Basics & Best Practices.
- Examples of different Table Extraction Pipelines.
- Practical problems & solutions.
- DocVQA as an alternative.
- Examples.

Summary and next steps

Next steps…

- We are adding new improved models all the time.
- New output formats for Tables.
- Trainable models.
- Want to try yourself? Ask for a trial!, enes@johnsnowlabs.com

Questions & Answers