

OCR and Document Understanding

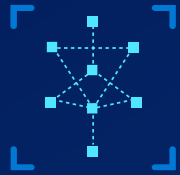


CHA ZHANG

MICROSOFT CLOUD & AI

Azure Cognitive Services

Pre-Trained and Customizable models with your data



Vision

Identify and analyze content within images, videos, and digital ink



Speech

Integrate speech processing into apps and services



Language

Extract meaning from unstructured text

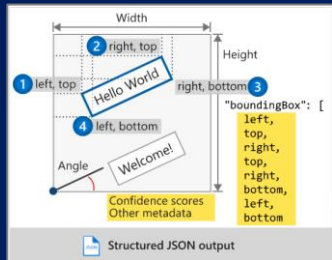


Decision

Make smarter decisions faster

As Part of Cognitive Service:

OCR (Read API)



OCR includes:

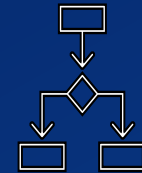
- Pages
- Text lines
- Words
- Locations



Unlocked text



Customers and partners add processing to get intelligent insights



Process automation

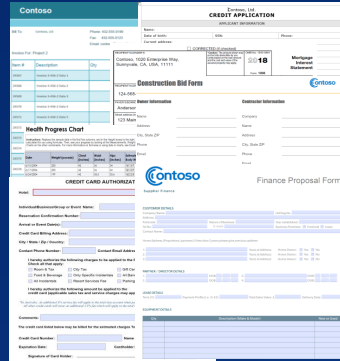
Document Understanding (Form Recognizer APIs)



Layout



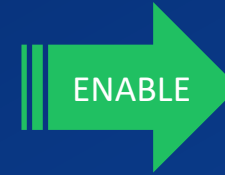
Pre-built



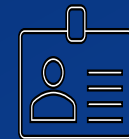
Custom

Form Recognizer includes:

- Text extraction
- Document structure
 - Tables
 - Selection marks
- Fields and values
- Other intelligence



Knowledge mining



Industry specific applications



OCR (Read API)

The thing I am concerned about, and so is Mr. Katzenbach, is having something issued so we can convince the public that Oswald is the real assassin. Mr. Katzenbach thinks that the President might appoint a Presidential Commission of three outstanding citizens to make a determination. I countered with a suggestion that we make an investigative report to the Attorney General with pictures, laboratory work, etc. Then the Attorney General can make the report to the President and the President can decide whether to make it public. I felt this was better because there are several aspects which would complicate our foreign relations. For instance, Oswald made a phone call to the Cuban Embassy in Mexico City which we intercepted. It was only about a visa, however. He also wrote a

S
E
R
E
T

The thing I am concerned about, and so is Mr. Katzenbach, is having something issued so we can convince the public that Oswald is the real assassin. Mr. Katzenbach thinks that the President might appoint a Presidential Commission of three outstanding citizens to make a determination. I countered with a suggestion that we make an investigative report to the Attorney General with pictures, laboratory work, etc. Then the Attorney General can make the report to the President and the President can decide whether to make it public. I felt this was better because there are several aspects which would complicate our foreign relations. For instance, Oswald made a phone call to the Cuban Embassy in Mexico City which we intercepted. It was only about a visa, however. He also wrote a

S
E
R
E
T

Challenges for “Universal OCR”

Large scale variability

Large aspect ratio

Cannot be enclosed tightly by axis-aligned rectangles

- e.g., skewed/curved text-lines

Nearby small-size text-lines

- e.g., inter-line space could be less than 2 pixels

Complex/ambiguous layout

Text-like background

- e.g., fences, bricks, stripes

High localization accuracy is required for text recognition engine

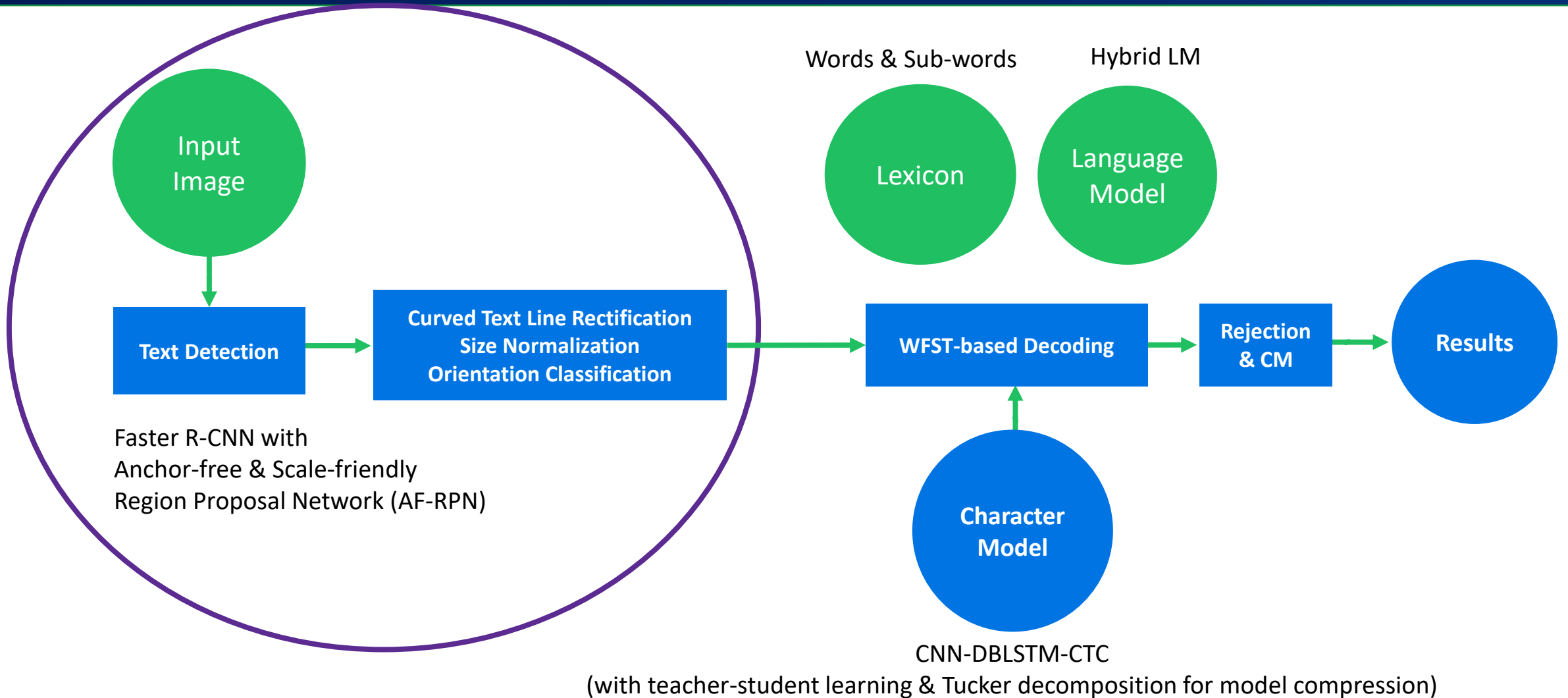
- “IOU>0.5” criterion is far from enough

Resolutions of input images cannot be reduced aggressively

- to avoid excessive small text instances

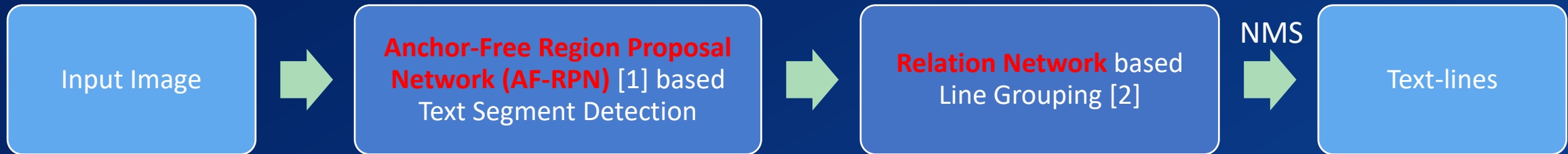


Read API: Microsoft's New Generation OCR Engine



A **unified** engine to recognize mixed printed and handwritten text lines with arbitrary orientations (even flipped)

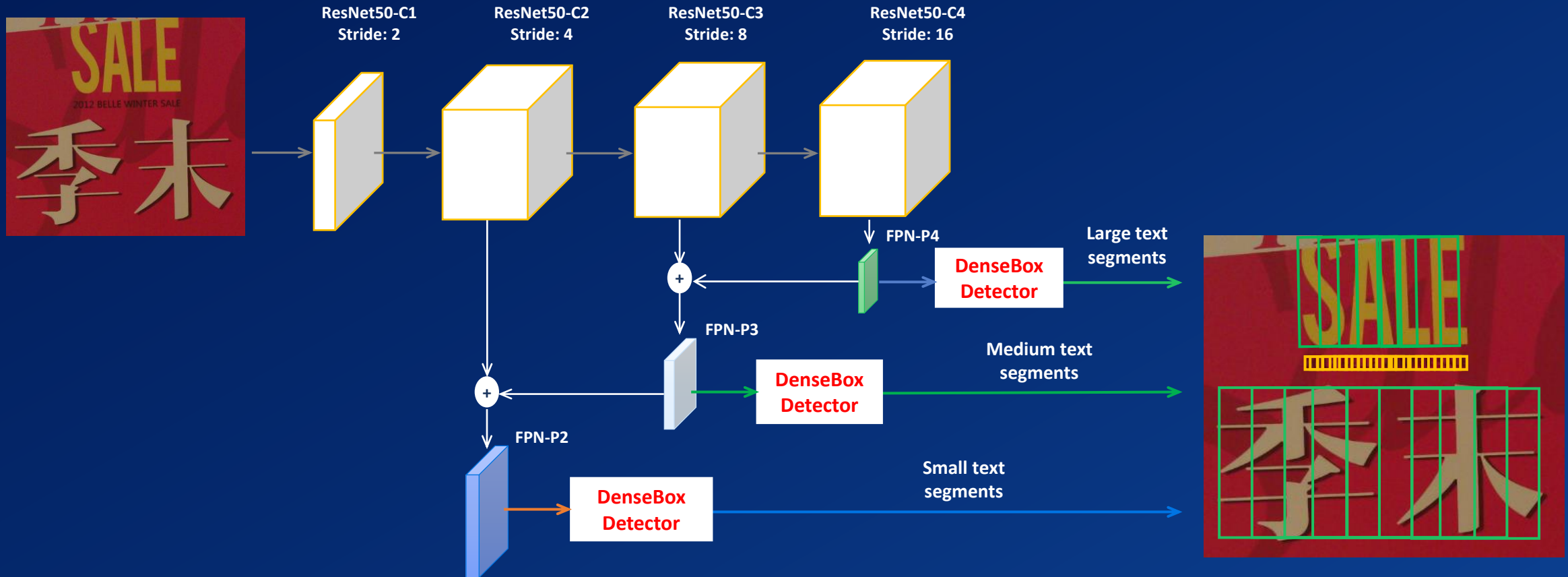
Our Text Detection Approach



[1] JS1-1 “An Anchor-Free Region Proposal Network for Faster R-CNN based Text Detection Approaches,”
Zhuoyao Zhong, Lei Sun, Qiang Huo, ICDAR-2019 oral presentation of the IJDAR paper

[2] PS2-07 “A Relation Network Based Approach to Curved Text Detection,”
Chixiang Ma, Zhuoyao Zhong, Lei Sun, Qiang Huo, ICDAR-2019

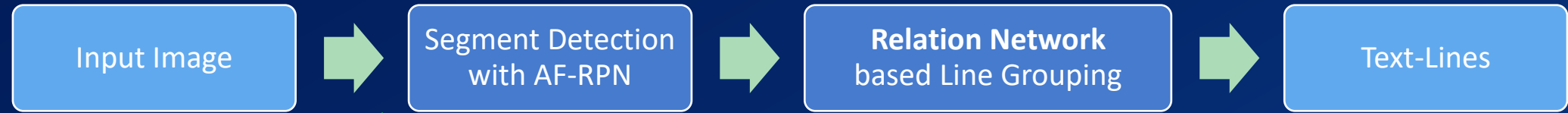
Anchor-Free Region Proposal Network (AF-RPN)



Scale-friendly learning: each DenseBox [1] only detects texts of scales within an appropriate range.

[1] L.-C. Huang, Y. Yang, Y.-F. Deng, and Y.-N. Yu, "Unifying landmark localization with end to end object detection," arXiv, 2015.

Relation Network based Line Grouping

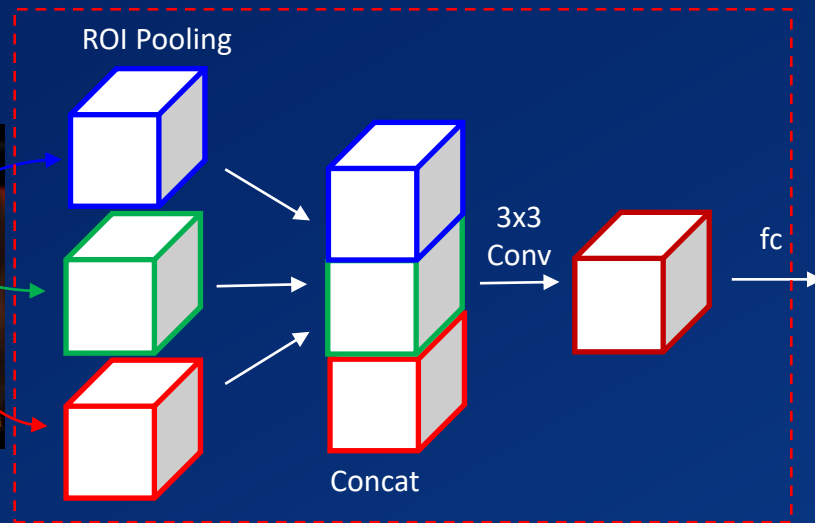


NMS

- Subject Segment
- Object Segment
- Relation BBox



Subject & Object Segment Pair Construction



Relation Network^[1]

Inter-Segment Link Relationship Prediction



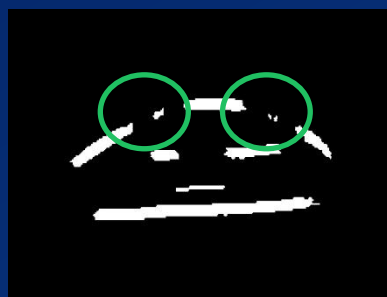
Line Grouping



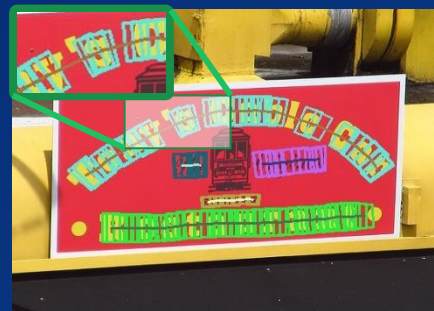
[1] J. Zhang, M. Elhoseiny, S. Cohen, W. Chang, and A. Elgammal, "Relationship proposal networks," in CVPR, 2017, pp. 5678-5686.

Advantages of Relation Network based Line Grouping

- Leverage the link between the pair of relatively distant segments
 - Able to detect text-lines with large inter-character spaces robustly
- Leverage wider context information to improve link prediction accuracy
 - More robust



Textness score map



Relation Network



SegLink vs. Relation Network

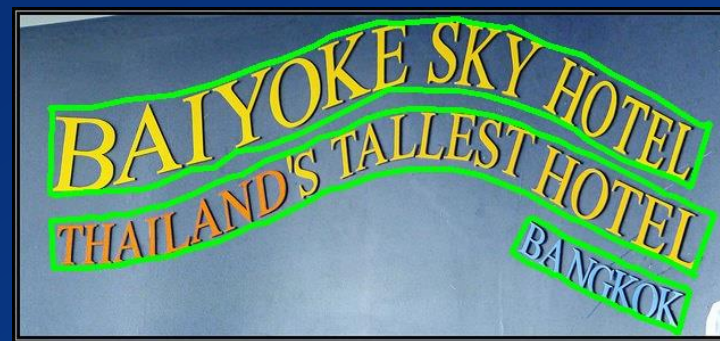


SegLink^[1]

Relation Network

[1] B.-G. Shi, et al., "Detecting oriented text in natural images by linking segments," CVPR, 2017.

Examples on SCUT-CTW1500^[1]



[1] Y.-L. Liu, L.-W. Jin, S.-T. Zhang, S. Zhang, "Detecting curve text in the wild: New dataset and new solution," arXiv, 2017.

Challenge of Detecting Small Text in High Resolution Images



Raw high-resolution image



Resized low-resolution image



Naive solution: use high-resolution image => Very high computation cost

How to detect small texts efficiently in high-resolution images?

Our Solution: Region-wise Adaptive Scaling

Input Image

1st Stage Text Block Detection and Scale Estimation

Region-wise Adaptive Scaling

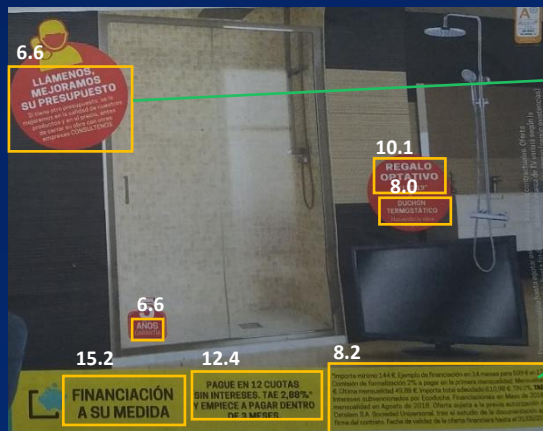
2nd Stage Text Detection

Text-lines

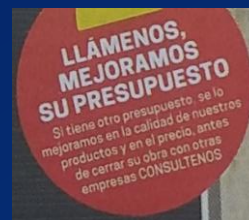
- Coarsely localize text-block regions from a low-resolution input image
- Resize each text-block to make shorter side lengths of contained texts in a working range



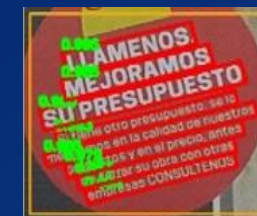
Resized low-resolution image



Detected text-blocks with estimated scales



Adaptively re-scaled text-blocks



Detected texts on each Re-scaled Text-block

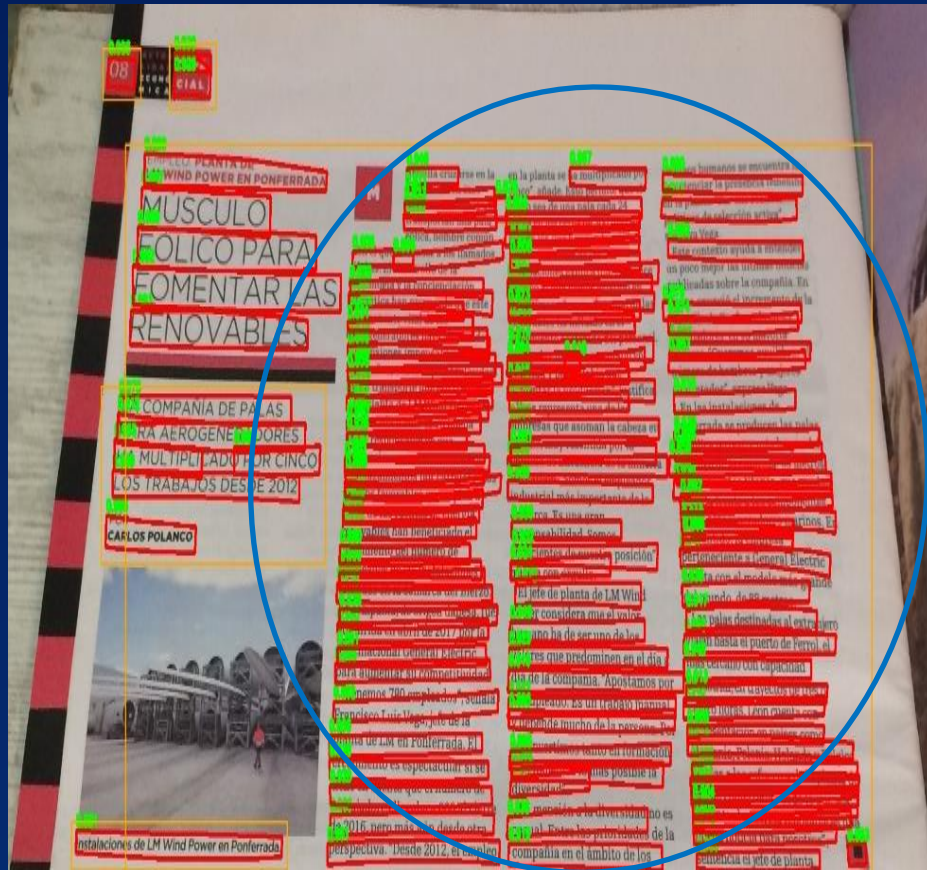


Final detected texts

Coarse

Fine

Example

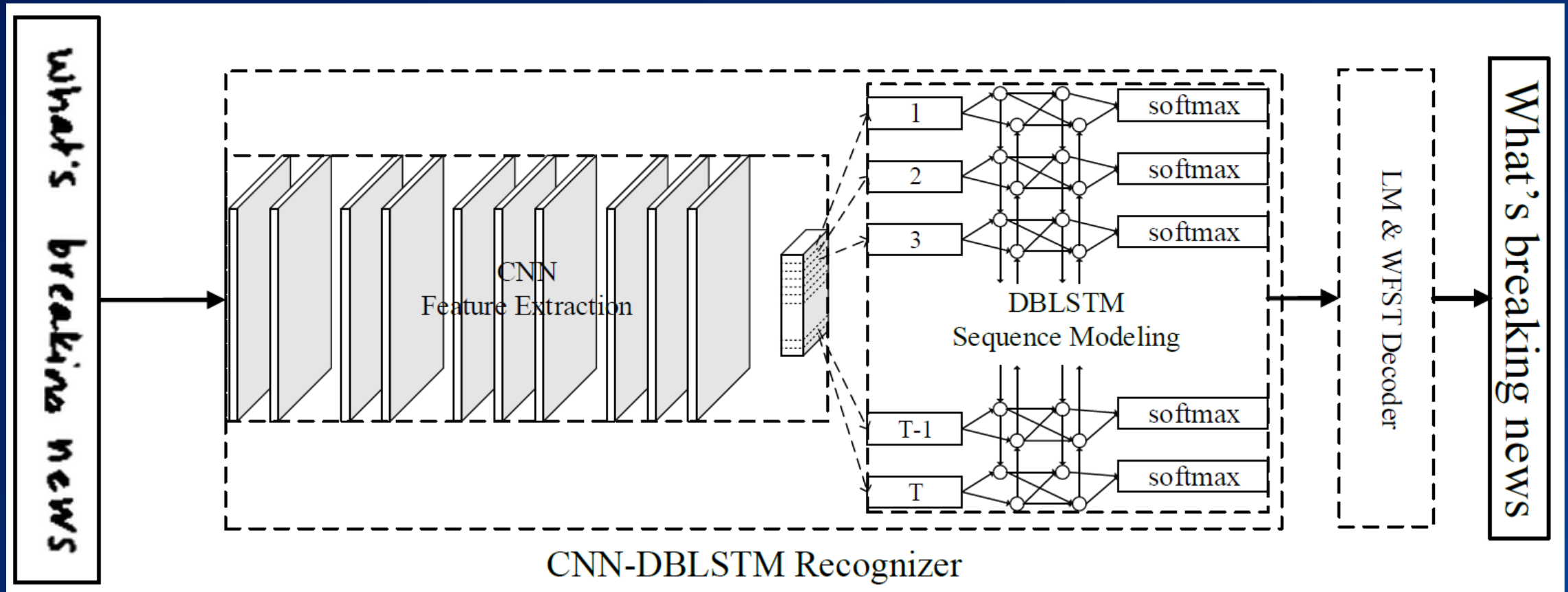


Without adaptive scaling



With region-wise adaptive scaling

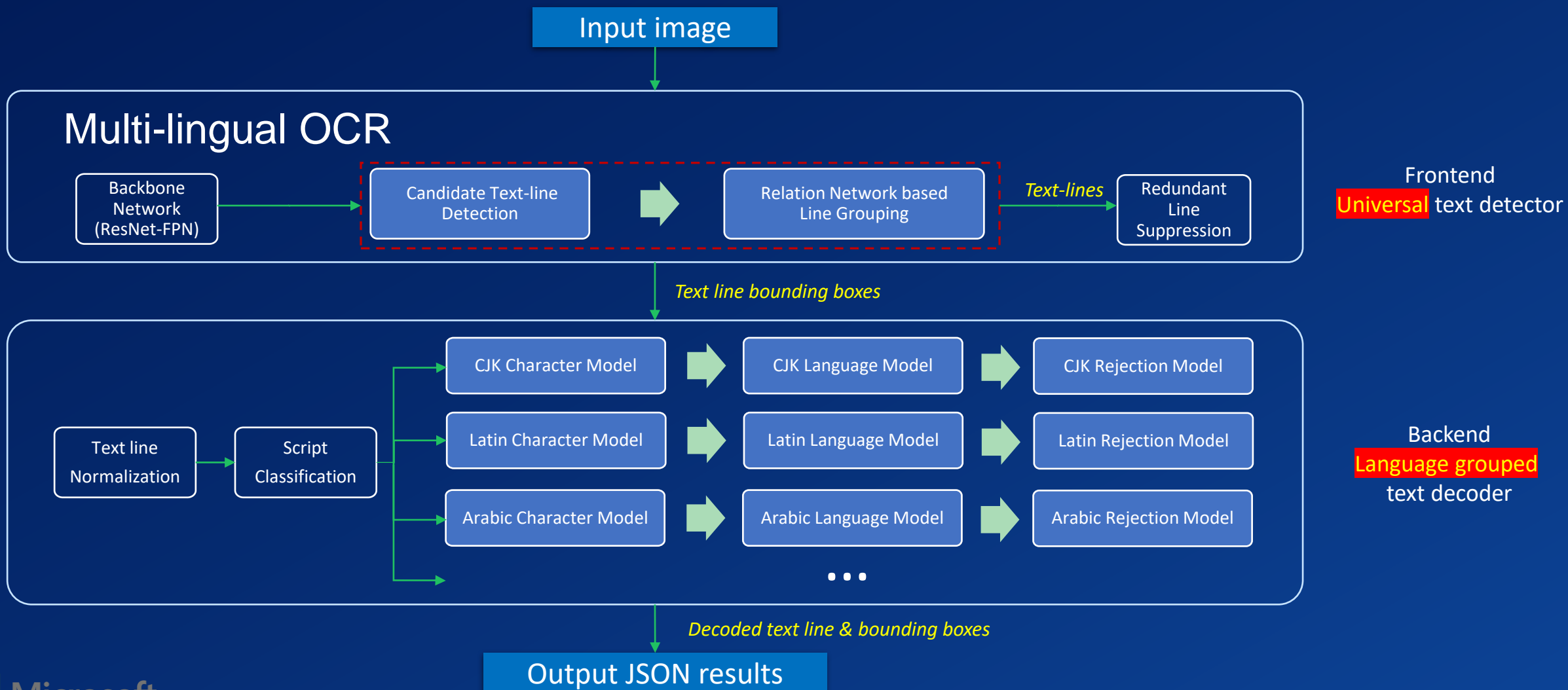
CNN-DBLSTM based Text Decoder



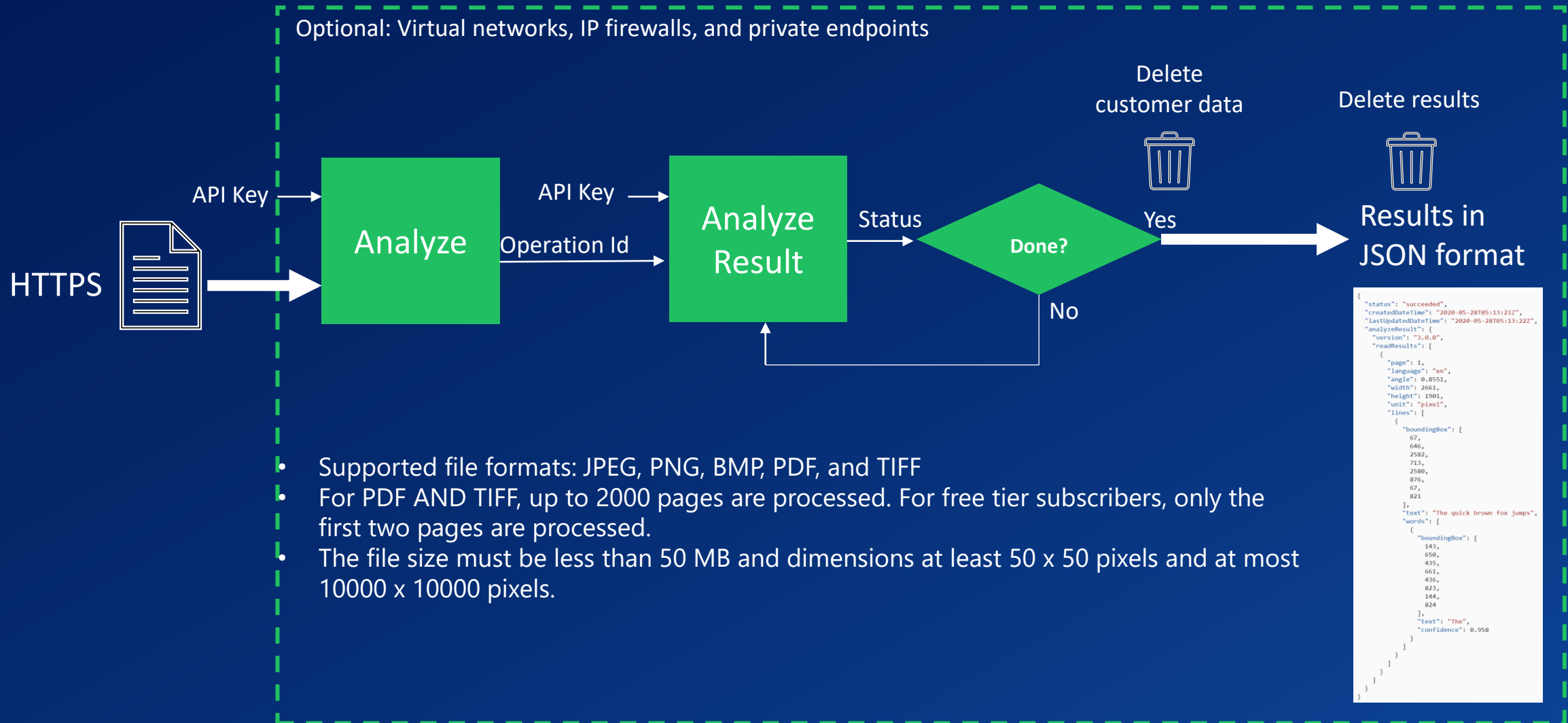
Input

WFST-based Decoder

Language Expansion



Read API – Available in Cloud and on-Prem



Mixed Languages Comparison (Example)



Other OCR auto mode



Microsoft OCR auto mode

Read 3.0+ Examples

1. Text in documents
2. Text in the wild
3. Languages

Indexed investing. The role and size of indexed investing continues to grow inexorably because it provides focused, efficient and transparent returns at relatively lower costs. Over the last 50 years, MSCI has provided indexes and other factors that have created new investment opportunities in various market sectors. We have advanced our research capabilities and created a new set of indices that have become the backbone of the global investment industry.

Recognized line: [relatively lower costs. Over the last 50 years, MSCI has provided indexes and other factors that have created new investment opportunities in various market sectors. We have advanced our research capabilities and created a new set of indices that have become the backbone of the global investment industry.]

boundingBox: 46,64,570,63,570,80,46,81

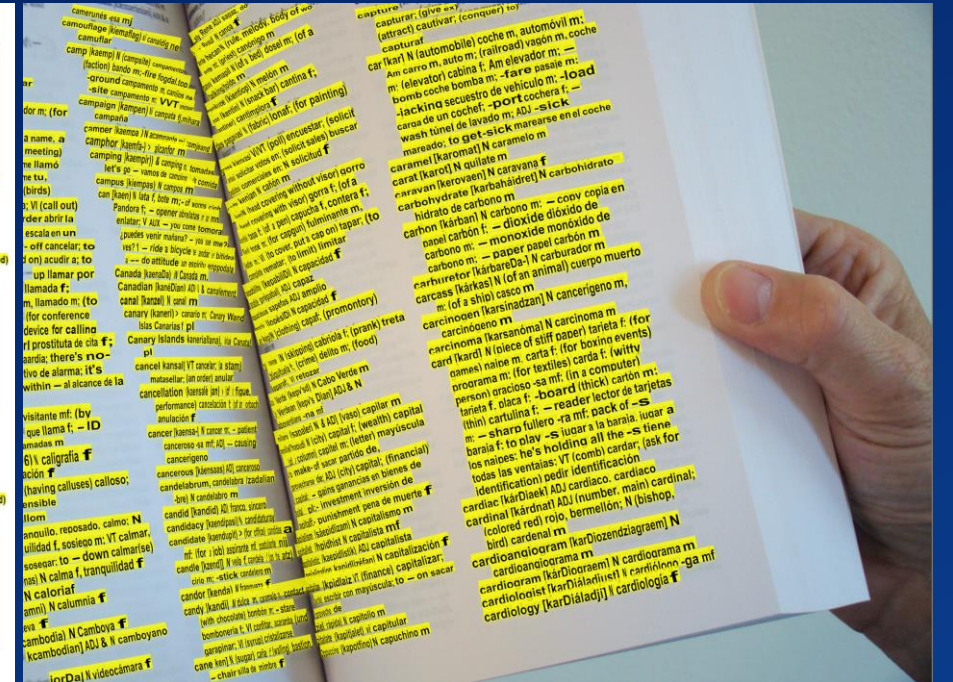
Factor investing. Barra (a part of MSCI) introduced the world to factor investing. Barra (a part of MSCI) introduced the world to factor investing. Barra (a part of MSCI) introduced the world to factor investing. Barra (a part of MSCI) introduced the world to factor investing. Barra (a part of MSCI) introduced the world to factor investing.

Recognized line: [>> Factor investing. Barra (a part of MSCI) introduced the world to factor investing.]

boundingBox: 21,437,526,437,526,455,21,456

MSCI's strategy

Our strategy is to deliver must-have, integrated, research-driven solutions to the owners and managers of capital around the world. The transparency provided by MSCI's offerings delivers insights and clarity into markets and assets that might have otherwise been opaque or unknown to investors. In an increasingly complex investment world, we offer solutions that provide greater understanding, transparency and clarity.



Images in the Wild



Documents

AUTOMATION ANYWHERE

Customer: Automation Anywhere
 Industry: Partner-Professional Services
 Size: 1,000 - 9,999 employees
 Country: United States

Products and services: Microsoft Azure, Azure Cognitive Services, Computer Vision, Form Recognizer, Optical Character Recognition

Read full story here

Microsoft

una guida rapida a ...
VE VENEZIA
 Non ci vorrà molto prima che ti innamorati di Venezia. La meravigliosa città d'acqua. Sperimentale sua sorprendente architettura e arte che si estende in modo seducente attraverso canali scintillanti e ponti decorati.

6/18/2020 Computer Vision for Augmented Reality - Microsoft Tech Community - 318099

Input: Length, Color, Port

Output: weight

importance) between
 supervised learning then?
 keting and sales, this is for 2.00m. Or that USB-C is
 rvised and active). For
 output. But what if the right combined with a certain

a shortcut guide to ...
VENICE
 It won't take long before you fall in love with Venice, the wonderful city of water. Experience its astonishing architecture and art that sprawl enticingly across sun-spangled canals and ornate bridges.

POPULAR MECHANICS

WIN

Microsoft's New HoloLens
 Personal services
 HoloLens
 Custom
 HoloLens

YOU CAN SEE THE
FUTURE
 WITH THESE

More
-26 BOLD IDEAS
 That Are Changing Our Times

MENU

STARTERS
 Add Text Here
 Add Text Here
 Add Text Here
 Add Text Here
 Add Text Here
 Add Text Here
 Add Text Here

SEAFOOD
 Add Text Here
 Add Text Here
 Add Text Here
 Add Text Here
 Add Text Here
 Add Text Here
 Add Text Here

CHEF'S SPECIALS
 Add Text Here
 Add Text Here
 Add Text Here
 Add Text Here

DESSERTS
 Add Text Here
 Add Text Here
 Add Text Here

pen and paper? I'd argue the biggest thing you lose is personality. Times New Roman looks the same on every paper, and frankly, it's boring to look at. Handwriting adds another dimension to your writing and make it truly yours. ■ can

TC# 6298 9647 9579 6887 9704
 NEED A MONEY ORDER? WAL-MART HAS IT!
 02/14/03 09:46:31
 *** CUSTOMER COPY ***

February
 March
 April
 May
 June
 July

GPU

HoloLens Graphics v

GPU engine utilization
 100.00

Frame rate
 120.00

App frames per second: 35.2

A HoloLens running RS4 (1803) is not able to use the GPU, all predictions will fail. If your device is upgraded to RS5 (1805) you are able to utilize the GPU output. There is one catch though: if you're using the CustomVision service it as v1.2 otherwise MLGen will give you a completely random error message

My colleague Sebastian described the Windows ML implementation in great

Documents...

MSCI World Index (USD)

The MSCI World Index captures large and mid cap representation across 23 Developed Markets (DM) countries*. With 1,601 constituents, the index covers approximately 85% of the free float-adjusted market capitalization in each country.

CUMULATIVE INDEX PERFORMANCE (JUL 2005 – JUL 2020)



ANNUAL PERFORMANCE (%)

Year	MSCI World	MSCI Emerging Markets	MSCI ACWI
2019	27.67	18.42	26.60
2018	-8.71	-14.57	-9.41
2017	22.40	37.28	23.97
2016	7.51	11.19	7.86
2015	-0.87	-14.92	-2.36
2014	4.94	-2.19	4.16
2013	26.68	-2.60	22.80
2012	15.83	18.22	16.13
2011	-5.54	-18.42	-7.36
2010	11.76	18.88	12.67
2009	29.99	78.51	34.63
2008	-40.71	-53.33	-42.19
2007	9.04	39.42	11.66
2006	20.07	32.14	20.95

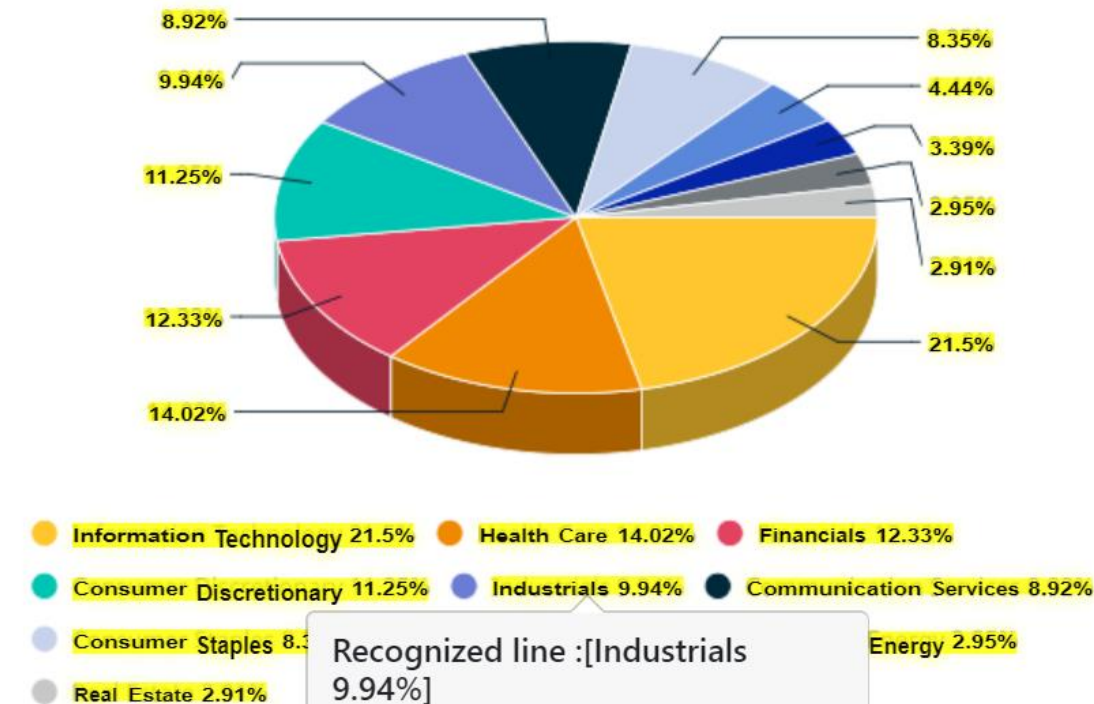
INDEX PERFORMANCE — NET RETURNS (%) (JUL 31, 2020)

	ANNUALIZED							
	1 Mo	3 Mo	1 Yr	YTD	3 Yr	5 Yr	10 Yr	Since Dec 29, 2000
MSCI World	4.78	12.75	7.23	-1.26	7.52	7.52	9.61	5.28

FUNDAMENTALS (JUL 31, 2020)

	Div Yld (%)	P/E	P/E Fwd	P/BV
MSCI World	2.08	21.56	20.57	2.55

SECTOR WEIGHTS

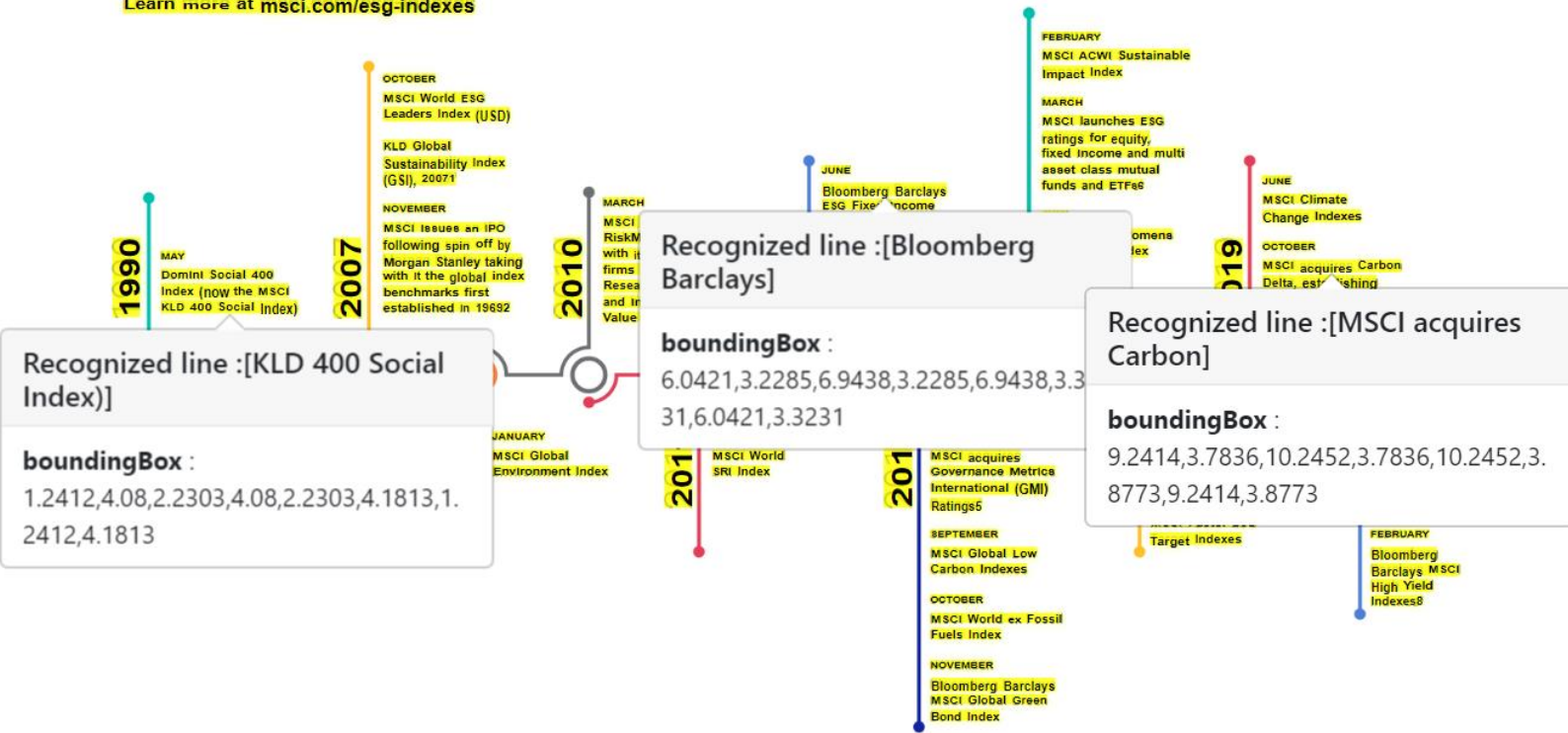


Visuals...

30 years of MSCI ESG Indexes

Take a look at the history of ESG indexes in our timeline below. We highlight key milestones in the evolution of ESG indexes since 1990, beginning with the launch of the Domini 400 Social Index (now the MSCI KLD 400 Social Index), through to the launch of the MSCI Fixed Income ESG Indexes in 2020. We also highlight significant developments such as MSCI's acquisition of Carbon Delta in 2019.

Learn more at [msci.com/esg-indexes](https://www.msci.com/esg-indexes)



1 <https://ir.msci.com/static-files/5556f636-a5e8-4377-8525-26a70587b106>
 2 <https://www.carmin.com/press-releases/23156-KLD-Launches-Q12018-Sustainability-Index-CSI>
 3 <https://ir.msci.com/static-files/15968432-4c08-424a-8e5c-479450273e50>
 4 https://www.msci.com/documents/10199/248121/Barclays_MSCI_ESG_Fixed_Income_Indexes_-_FINAL.pdf?00f030-1573-46ad-4725-050790059a87?version=1.0
 5 <https://ir.msci.com/static-files/12c28970-c8a0-45db-8565-83781144b81e>
 6 <https://www.msci.com/www/blog-posts/msci-introduces-esg-quality-0308840040>
 7 <https://ir.msci.com/news-releases/news-release-details/msci-strengthens-climate-risk-capability-acquisition-carbon-delta>
 8 <https://www.bloomberg.com/press-releases/2020-02-11/bloomberg-and-msci-expand-esg-fixed-income-index-suite-launch-new-esg-high-yield-indexes>

© 2020 MSCI Inc. All rights reserved.



Research Insights



China and the Future of Equity Allocations

What does the partial inclusion of China A shares in MSCI indexes mean for global and emerging market equity portfolios?

[Learn More](#)



Liquidity and Correlation in the Chinese Credit Market

China's stock market has drawn huge attention from global investors.

Recognized line :[China's stock market has drawn huge]

boundingBox :
 321,456,527,456,527,472,321,471



The Rise of Fundamental Factors in China A Shares

Commonly held perceptions about China A shares have influenced investors to think factor strategies work in the Chinese equity market. Our research suggests this perception is changing.

[Learn More](#)



Stress Testing US-China Trade Wars

Amid ongoing U.S.-China trade tensions, we have updated our stress test to consider three scenarios for how the situation could unfold—and their impact on currency, bond and equity markets around the world.

[Learn More](#)



Chinese Convertibles: Equities in Fancy Dress?

Chinese corporate bonds that convert to A shares deserve equity-like

Recognized line :[Chinese corporate bonds that convert]

boundingBox :
 321,939,530,938,530,952,321,953



China Through an ESG Lens

Chinese domestic investors and issuers are moving fast to incorporate ESG considerations in their decision making, propelled by regulatory initiatives to promote ESG practices and disclosure. At the same time, shortages of skilled talent, consumer expectations around safety, growing climate risks and increased regulations for shareholder rights

[Learn More](#)

Mixed Languages...

FOOD . LA NOURRITURE . DIE NAHRUNGSMITTEL . LOS ALIMENTOS . IL CIBO

nuts and dried fruit . les noix et les fruits secs . die Nüsse und das Dörrobst
 los frutos secos - le noci e la frutta secca

 pine nut . le pignon die Piniennuss el piñón . il pinolo	 pistachio . la pistache die Pistazie . el pistacho . il pistacchio	 cashewnut . la noix de cajou . die Cashewnuss . el anacardo . l'anacardio	 peanut . la cacahuète . die Erdnuss . el cacahuete l'arachide	 hazelnut . la noisette die Haselnuss la avellana la nocciola
 brazilnut . la noix du Brésil . die Paranuss la nuez de Brasil la mandorla brasiliana	 pecan . la noix pacane . die Pecannuss la pacana . la noce pecan	 almond . l'amande die Mandel . la almendra . la mandorla	 walnut . la noix . die Walnuss . la nuez la noce	 chestnut . le marron die Esskastanie . la castaña . la castagna
 macadamia le Macadamia die Macadamianuss la macadamia la noce di macadamia	 fig . la figue . die Feige . el higo . el fico	 date . la datte die Dattel . el dátíl il dattero	 prune . le pruneau die Backpflaume la ciruela pasa la prugna secca	 shell la coquille die Schale la cáscara il guscio
 sultana . le raisin de Smyrne . die Sultanine la pasa sultana	 raisin . le raisin sec die Rosine . la pasa l'uvetta	 currant . le raisin de Corinthe . die Korinthe la pasa de Corinto	 coconut . la noix de coco . die Kokosnuss el coco . la noce	 flesh la chair das Fruchtfleisch la pulpa la polpa

城市大脑
City Brain: A City Intelligence Infrastructure

定义

- 城市治理的前瞻性探索实践
- 全局实时分析城市运行状态
- 利用数据资源调配公共资源
- 修正城市运行过程中的缺陷

阿里云 奥运会全球和宝营销服务商

SUMMER FRUIT 鲜榨果汁 (400cc/500cc)	西瓜汁 梨汁 橙汁 西瓜+梨 梨+橙 西瓜+橙 梨+柠檬 橙+柠檬	FRUIT TEA 霸煮水果茶 (1000cc)	霸气水果茶 霸气芒果 霸气柠檬 霸气橙子 霸气红柚
CLASSIC PORT 经典港式 (400cc)	丝袜奶茶 丝袜奶盖 红豆奶茶 血糯米奶茶 珍珠奶茶 港式咖啡 香港鸳鸯 鸳鸯奶盖	SUMMER TEA 夏日凉饮 (500cc/1000cc)	冻柠茶 柠檬水 冻柠蜜 冻柠七 龙凤柠七
CLASSIC FOOD 港式小吃	原味鸡蛋仔 咸蛋黄蛋仔 巧克力蛋仔 葡萄干蛋仔 黑芝麻蛋仔 桂花鸡蛋仔 红豆鸡蛋仔	CLASSIC FOOD 港式小吃	咖喱鱼丸 咖喱鱼蛋 咖喱双拼

	Drive on the left
	Conduise à gauche
	Links fahren
	Tenere la sinistra
	Conduzca por la izquierda
	Links rijden

www.LEGOsurvey.com/pr

GEWINNE! **WIN!** **GAGNE!**

Nimm an der Umfrage auf www.LEGOsurvey.com/product teil und hab die Chance ein cooles LEGO® Produkt zu gewinnen!
 Die Teilnahme am Gewinnspiel führt zu keinem Kaufverpflichtungen. Teilnahme in allen nicht-beschlossenen Ländern möglich.

Ga naar www.LEGOsurvey.com/product voor een enquêteformulier in en maak kans op een cool LEGO® product.
 Geen aankoopverplichting. Iedereen mag deelnemen, uitzonderend in bepaalde landen waar een enquêteverbod geldt.

Visite www.LEGOsurvey.com/product pour répondre à un questionnaire et avoir une chance de gagner un produit LEGO® très cool!
 Aucune obligation d'achat. L'événement est ouvert à tous les résidents des pays autorisés.

LEGO and the LEGO logo are trademarks of the LEGO Group. ©2011



Document Understanding (Form Recognizer API)

CONTOSO LTD.

Contoso Headquarters
123 456th St
New York, NY, 10001

Microsoft Corp
123 Other St,
Redmond WA, 98052

Microsoft Finance
123 Bill St,
Redmond WA, 98052

REMIT TO:
Contoso Billing
123 Remit St
New York, NY, 10001

INVOICE

INVOICE: INV-100
INVOICE DATE: 11/15/2019
DUE DATE: 12/15/2019
CUSTOMER NAME: MICROSOFT CORPORATION
SERVICE PERIOD: 10/14/2019 – 11/14/2019
CUSTOMER ID: CID-12345

SHIP TO: Microsoft Delivery
123 Ship St,
Redmond WA, 98052

SERVICE ADDRESS: Microsoft Services
123 Service St,
Redmond WA, 98052

SALESPERSON	P.O. NUMBER	REQUISITIONER	SHIPPED VIA	F.O.B. POINT	TERMS
	PO-3333				

QUANTITY	DESCRIPTION	UNIT PRICE	TOTAL
1	Consulting service	1	\$100.00

SUBTOTAL \$100.00
SALES TAX \$10.00
TOTAL \$110.00
PREVIOUS UNPAID BALANCE \$500.00
TOTAL DUE \$610.00

THANK YOU FOR YOUR BUSINESS!

Page # / Field name / Value	Confidence
1 AmountDue text: \$610.00 valueNumber: 610	88.20%
1 BillingAddress 123 Bill St, Redmond WA, 98052	99.70%
1 BillingAddressRecipient Microsoft Finance	99.80%
1 CustomerAddress 123 Other St, Redmond WA, 98052	99.80%
1 CustomerAddressRecipient Microsoft Corp	99.70%
1 CustomerId CID-12345	98.00%
1 CustomerName MICROSOFT CORPORATION	98.20%
1 DueDate text: 12/15/2019 valueDate: 2019-12-15	99.50%
1 InvoiceDate text: 11/15/2019 valueDate: 2019-11-15	99.60%
1 InvoiceId INV-100	99.90%
1 InvoiceTotal text: \$110.00 valueNumber: 110	98.90%
1 PreviousUnpaidBalance text: \$500.00 valueNumber: 500	98.90%
1 PurchaseOrder PO-3333	96.10%

```
"documentResults": [{"docType": "Invoice", "pageRange": "1-1"}, {"fields": {"AmountDue": {"type": "Text", "value": "$610.00", "text": "$610.00", "boundingBox": "7.81 7.91 7.81 7.91 7.95 7.38 7.95 7.95"}, "BillingAddress": {"type": "Text", "value": "123 Bill St, Redmond WA, 98052", "text": "123 Bill St, Redmond WA, 98052", "boundingBox": "4.37 2.01 4.37 2.01 4.71 0.59 4.71 0.59"}, "BillingAddressRecipient": {"type": "Text", "value": "Microsoft Finance", "text": "Microsoft Finance", "boundingBox": "0.59 4.71 0.59 4.71"}, "CustomerId": {"type": "Text", "value": "CID-12345", "text": "CID-12345", "boundingBox": "0.59 4.71 0.59 4.71"}, "CustomerName": {"type": "Text", "value": "MICROSOFT CORPORATION", "text": "MICROSOFT CORPORATION", "boundingBox": "0.59 4.71 0.59 4.71"}, "DueDate": {"type": "Text", "value": "2019-12-15", "text": "12/15/2019", "boundingBox": "0.59 4.71 0.59 4.71"}, "InvoiceDate": {"type": "Text", "value": "2019-11-15", "text": "11/15/2019", "boundingBox": "0.59 4.71 0.59 4.71"}, "InvoiceId": {"type": "Text", "value": "INV-100", "text": "INV-100", "boundingBox": "0.59 4.71 0.59 4.71"}, "InvoiceTotal": {"type": "Text", "value": "110", "text": "$110.00", "boundingBox": "0.59 4.71 0.59 4.71"}, "PreviousUnpaidBalance": {"type": "Text", "value": "500", "text": "$500.00", "boundingBox": "0.59 4.71 0.59 4.71"}, "PurchaseOrder": {"type": "Text", "value": "PO-3333", "text": "PO-3333", "boundingBox": "0.59 4.71 0.59 4.71"}}, "page": "1"}]
```

Form Recognizer

Data extraction in any business process that intakes forms and outputs structured data



Layout



Prebuilt[s]

Item #	Description	City
20087	Invoice 3-458-2 Data 1	
20088	Invoice 3-458-2 Data 2	
20089	Invoice 3-458-2 Data 3	
20090	Invoice 3-458-2 Data 4	
20091	Invoice 3-458-2 Data 5	

Date	Weight (pounds)	Chest (inches)	Waist (inches)	Hips (inches)	Actual Body Fat
11/10/2004	200	40	34	34	14.15%
11/17/2004	200	40	34	34	14.15%
11/24/2004	199	40	34	34	14.15%

Custom

An Atypical Language Understanding Problem

租税条約に関する届出書
APPLICATION FORM FOR INCOME TAX CONVENTION

Recognized line :[Relief from Japanese Income Tax and Special Income]

boundingBox : 3.2059,1.1748,5.8717,1.1748,5.8717,1.26
89,3.2059,1.2689

appearance.style : print
appearance.styleConfidence : 1

Recognized line :[上記「3」の支払者から支払を受ける配当で「1」の租税条約の規定の適用を受けるものに関する事項（注10）；]

boundingBox : 0.8577,8.7963,6.4117,8.7963,6.4117,8.89
66,0.8577,8.8966

appearance.style : print
appearance.styleConfidence : 1

元本の種類	銘柄	数量	議決権のある株式数	元本の取得年月日	
Kind of Principal	Description	Name of Nominee of Principal (Note 11)	Quantity of Principal	Of which Quantity of Voting Shares	Date of Acquisition of Principal
<input type="checkbox"/> 出資株式基金 Shares (Stocks)					
<input type="checkbox"/> 株式投資信託 Stock investment trust					

Not all documents can have a clear read order... Can we still extract knowledge like key-value pairs?

Visual Linguistic Tasks

Visual Question Answering

Who is wearing glasses?
man woman



Where is the child sitting?
fridge arms



Is the umbrella upside down?
yes no



How many children are in the bed?
2 1




Image Captioning



Image-Text Retrieval

Query: A man riding a motorcycle is performing a trick at a track .




Query: Two dogs play by a tree .



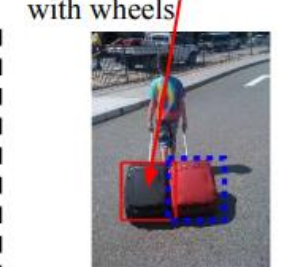

1:A female runner dressed in blue athletic wear is running in a competition , while spectators line the street . ✓
2:A lady dressed in blue running a marathon . ✓
3:A young woman is running a marathon in a light blue tank top and spandex shorts . ✓
4:A lady standing at a crosswalk . ✗
5:A woman who is running , with blue shorts . ✓

Referred Expression Comprehension

A black carry-on suitcase with wheels



The truck in the background.

Visual Commonsense Reasoning



Why is [person4] pointing at [person1]?

a) He is telling [person3] that [person1] ordered the pancakes.

b) He just told a joke.

c) He is feeling accusatory towards [person1].

d) He is giving [person1] directions.

Some Representative Works on Visual-Linguistic Joint Modeling

- VideoBERT: A Joint Model for Video and Language Representation Learning, Chen Sun, Austin Myers, Carl Vondrick, Kevin Murphy, Cordelia Schmid, ICCV 2019.
- ViLBERT: Pretraining Task-Agnostic Visiolinguistic Representations for Vision-and-Language Tasks, Jiasen Lu, Dhruv Batra, Devi Parikh, Stefan Lee, NIPS 2019.
- LXMERT: Learning Cross-Modality Encoder Representations from Transformers, Hao Tan, Mohit Bansal, EMNLP 2019.
- Unicoder-VL: A Universal Encoder for Vision and Language by Cross-modal Pre-training, Gen Li, Nan Duan, Yuejian Fang, Ming Gong, Daxin Jiang, Ming Zhou, AAAI 2020.
- VL-BERT: Pre-training of Generic Visual-Linguistic Representations, Weijie Su, Xizhou Zhu, Yue Cao, Bin Li, Lewei Lu, Furu Wei, Jifeng Dai, ICLR 2020.

Focusing More on Documents

- Yang et al.^[1] presented an end-to-end, multimodal, fully convolutional network for extracting semantic structures.
- Liu et al.^[2] introduced a Graph Convolutional Networks (GCN) based model to combine textual and visual information.
- Davis et al. ^[3] proposed to use relationship classifier and neighbor prediction networks to identify key-value pairs.
- Sarkhel et al. ^[4] proposed a multi-scale classification method to classify the visually rich document.

[1] Yang, Xiaowei et al. "Learning to Extract Semantic Structure from Documents Using Multimodal Fully Convolutional Neural Networks." CVPR (2017).

[2] Liu, Xiaojing et al. "Graph Convolution for Multimodal Information Extraction from Visually Rich Documents." NAACL-HLT (2019).

[3] Davis, Brian et al. "Deep Visual Template-Free Form Parsing." ICDAR (2019).

[4] Sarkhel, Ritesh and Arnab Nandi. "Deterministic Routing between Layout Abstractions for Multi-Scale Classification of Visually Rich Documents." IJCAI (2019).

Report

Form

Receipt

Invoice

UBS Global Research 31 January 2018

First Read

Microsoft Corp.

Azure Acceleration Impresses in Solid Q2

Good Q2 With More to Come
Microsoft posted a clean beat with revenue, margins, and EPS all topping Street expectations and H218 estimates moving higher. With shares +11% YTD, some of the good news is likely already priced in, but given the backdrop of accelerating Azure growth, an improving outlook for margins, and the potential for increased IT spending on the back of tax reform, we see more beats ahead and continue to recommend the stock while raising our price target from \$105 to \$110 on the back of our FY18 and FY19 estimates moving higher.

Azure Acceleration
Azure revenue growth of 98% was the strongest in 5 Qs, while Azure margins continued to show strong improvement YTD. Overall Cloud GM did disappoint slightly QoQ, to 55% from 57% in Q1 due to seasonality impacts linked to the timing of ongoing payments for some Azure contracts as well as Azure comprising a larger piece of revenue mix, but the YTD trend in Cloud GM remains healthy. Microsoft added \$1.2 of new Azure rev. YTD for each \$ of added IC. COGS, up from -\$1.55 in the prior 2 Qs and highlighting expanding incremental margins in the Azure business.

Revenue Beat and Tax Benefits Drive Estimates Higher
Gross and operating margin visible in Q2 drives margin expectations higher for the year, with the company now expecting flat GM YTD and a slight increase in GM for FY18. Our FY18 rev. est. moves from \$110.6B to \$107.2B while FY19 goes from \$116.0B to \$117.3B. Microsoft now expects +16% tax rate for H218 and just under 21% in FY19 and beyond vs. a prior rate of 23%. As a result, our FY18 EPS estimate moves to \$3.62 (\$3.35 prior) while FY19 moves to \$4.02 (\$3.78 prior).

Valuation: PT moves 5% higher to \$110 (\$105 prior) on positive revisions
We value Microsoft's Cloud businesses using a 5x multiple, which moves from 6.2x EV/EBITDA to 6.8x due to better growth and better FCF margins in the Cloud business, as well as broader multiple expansion across the SaaS group, while better cash flow also increases our valuation for the legacy on-premise businesses which we value assuming 3% annual decline in perpetuity with a 5% discount rate.

Highlights (\$US)	06/15	06/16	06/17	06/18	06/19	06/20	06/21	06/22
Revenues	93,380	91,154	96,571	107,960	117,341	129,386	142,726	157,148
EBIT (US\$)	28,172	27,188	29,311	32,724	36,812	45,469	52,003	61,117
Net earnings (US\$)	21,627	21,424	25,732	28,301	31,850	36,943	42,868	49,383
EPS (US\$ diluted)	2.62	2.68	3.29	3.62	4.02	4.87	5.43	6.25
DP (US\$)	2.21	2.44	1.53	1.62	1.77	1.80	1.80	1.80
Net debt/cash	61,234	59,779	66,787	81,794	106,868	138,852	176,724	226,527

www.ubs.com/investmentresearch

This report has been prepared by UBS Securities LLC. ANALYST CERTIFICATION AND REQUIRED DISCLOSURES BEGIN ON PAGE 3. UBS does and seeks to do business with companies covered in its research reports. As a result, investors should be aware that the firm may have a conflict of interest that could affect the objectivity of this report. Investors should consider this report as only a single factor in making their investment decision.

ACUTE TOXICITY IN MICE

COMPOUND: **3-Hydroxy-3-methylbutanoic acid (Tur 13)**

SOURCE: **Lorillard - Organic Chemistry** (LORILLARD NO. **OR39-23**)

DATE RECEIVED: **Unk.** TESTED: **12/28/78** REPORTED: **10/6/80, Update**

INVESTIGATOR: **H. S. Tonn & M. S. Forte** NOTEBOOK PAGE: **B1014-23**

SIGNATURES: *H.S. Tonn* *M.S. Forte (by G. Poole)*

STRAIN OF MICE: **Swiss-Webster** MALE FEMALE DATE RECEIVED: **Unk.**

AVERAGE WEIGHT/RANGE (GM): **Swiss Research**

ROUTE OF COMPOUND ADMINISTRATION: P.O. I.P. I.V. INHALATION

COMPOUND VEHICLE: 5% METHYL CELLULOSE CORN OIL SALINE OTHER

GROUP NO.	% SOLUTION	DOSE (mg/kg BODY WEIGHT)	RESULTS (NO. DEAD/NO. TESTED)
1	5	1800	1/6
2	10	2160	0/6
3	10	2520	0/6
4	10	3732	3/6
5	10	4479	6/6

REFERENCE FOR CALCULATION: **Litchfield, J. T. and Wilcoxon, F., J. of Pharmacol. and Exper. Ther., 90:99, 1948.**

LD50 (95% CONFIDENCE LIMITS): **3.5 (3.1 to 3.9) g/kg**

CONCLUSION: **This compound appears to act as a CNS depressant with symptoms of respiratory depression, constriction of blood vessels, and in-activity. Survivors recovered in 48 hours. The recommended safe dose for a single trial by inhalation in man is 0.3 mg.**

Copies to the Following: **Dr. H. J. Minnemeyer, Ms. L. B. Gray**

LORILLARD RESEARCH CENTER FORM 7 (5-80)

Morton's The Steakhouse
735 S Figueroa Street
Los Angeles, CA 90017
(213) 553-4566

Server: Sally DBB: 09/08/2016
06:23 PM 09/08/2016
404/1 1/10084

SALE 6291458

Visa Card #XXXXXXXXXX8698
Magnetic card present: SICKAFOOSE/DANNY
Card Entry Method: S

Approval: 047028

Amount: \$33.79
+ Gratuity Not Inc: 6
= Balance Due: 39.79

I agree to pay the above total amount according to the card issuer agreement.

For banquet events, balance due includes suggested gratuity if accepted.

Guest Copy

Access Information Protected. 10445 4th Avenue Denver, CO 80238

Page 1 of 1
Invoice
1 877 FileLine | InformationProtected.com

New Belgium Brewery Company
Attn: Accounts Payable Manager
500 Linden St
Ft Collins, CO 80524

Service Billing Period: 1/12/2017
Date: 1/12/2017
Invoice #: 1861619
Customer #: 01P9286

Total Amount Due: \$546.89
By: 12/28/17
Total Enclosed:

Remit To: PO Box 398303 San Francisco, CA 94139-8303
When making payment, please reference invoice number 1861619

NOTE: MAIN

QTY	ITEMS	SERVICE DESCRIPTION	QUANTITY	RATE	TAX	SEE
Storage						
Storage Period: 02/15/2017 - 02/28/2017						
4	Legal Bankers Box		10.00	0.5040	N	5.04
408	Letter Bankers Box		236.00	0.5040	N	471.75
85	Letter Legal Bank		85.00	0.5040	N	45.90
		TOTAL FOR Storage	1,031.00			622.89
		TAX				6.00
Service						
		File Tracking	3.00	0.0000	N	0.00
		Medium console - Initial Delivery	3.00	0.0000	N	0.00
		Medium Console - Scheduled Rotation / Plant	3.00	0.0000	N	0.00
		Container Refill	4.00	6.0000	N	24.00
		FILEDRIDGE Records + AccessMETRICS	1.00	0.0000	N	0.00
		TOTAL FOR Service				24.00
		TAX				6.00
Transportation						
		Shred Rotation Transportation - Scheduled trip	2.00	0.0000	N	0.00
		TOTAL FOR Transportation				6.00
		TAX				6.00
		SUB-TOTAL				546.89
		TAX				6.00
		INVOICE TOTAL				5546.89

PLEASE NOTE: To the extent you do not have a currently effective written contract for services with an Access or Retrievex company, by paying this invoice, you agree that the terms and conditions found on www.informationprotected.com/access-service-terms-and-conditions (December 1, 2016 version) will apply to and govern the storage, document destruction, imaging and other services provided to you by such company and, therefore, WILL AFFECT YOUR LEGAL RIGHTS AND OBLIGATIONS, AND LIMITS OUR LIABILITY TO YOU. However, if you have a currently effective written contract for services with an Access or Retrievex company, the terms and conditions of your written contract will continue to apply as provided in such contract. Further, if you are a "Covered Entity" or "Business Associate" as defined in 45 CFR part 160 and do not have a currently effective written Business Associate Agreement (BAA) or Business Associate Subcontractor Agreement (BASAs) with an Access or Retrievex company, by paying this invoice, you agree that the terms and conditions found on www.informationprotected.com/baa constitute a legally effective BAA or BASA as applicable, between you and such Access or Retrievex company. As determined appropriate by Access, payments that do not reference a specific invoice will be applied to the oldest outstanding invoice. Terms or conditions on purchase orders or similar documents submitted to Access or Retrievex are not binding.

Document Understanding in Real World

LayoutLM: Pre-training for Text with rich Layout and Style information [1]



[1] Xu et al., LayoutLM: Pre-training of Text and Layout for Document Understanding, KDD 2020.

Example: Invoice Understanding

Vendor Name: OMNI PROMOTIONAL, LLC
Vendor Address: 1555 CHERRY STREET, LOUISVILLE, CO 80027
Invoice Number: 0047076-IN
Invoice Date: 3/24/2017
Due Date: 4/25/2017

Customer Name: NEW BELGIUM BREWING
Customer Address: 500 LINDEN ST, FORT COLLINS, CO 80521, United States
Shipping Address: NBB - RESERVOR, 3620 WEICKER DR, FORT COLLINS, CO 80524, United States

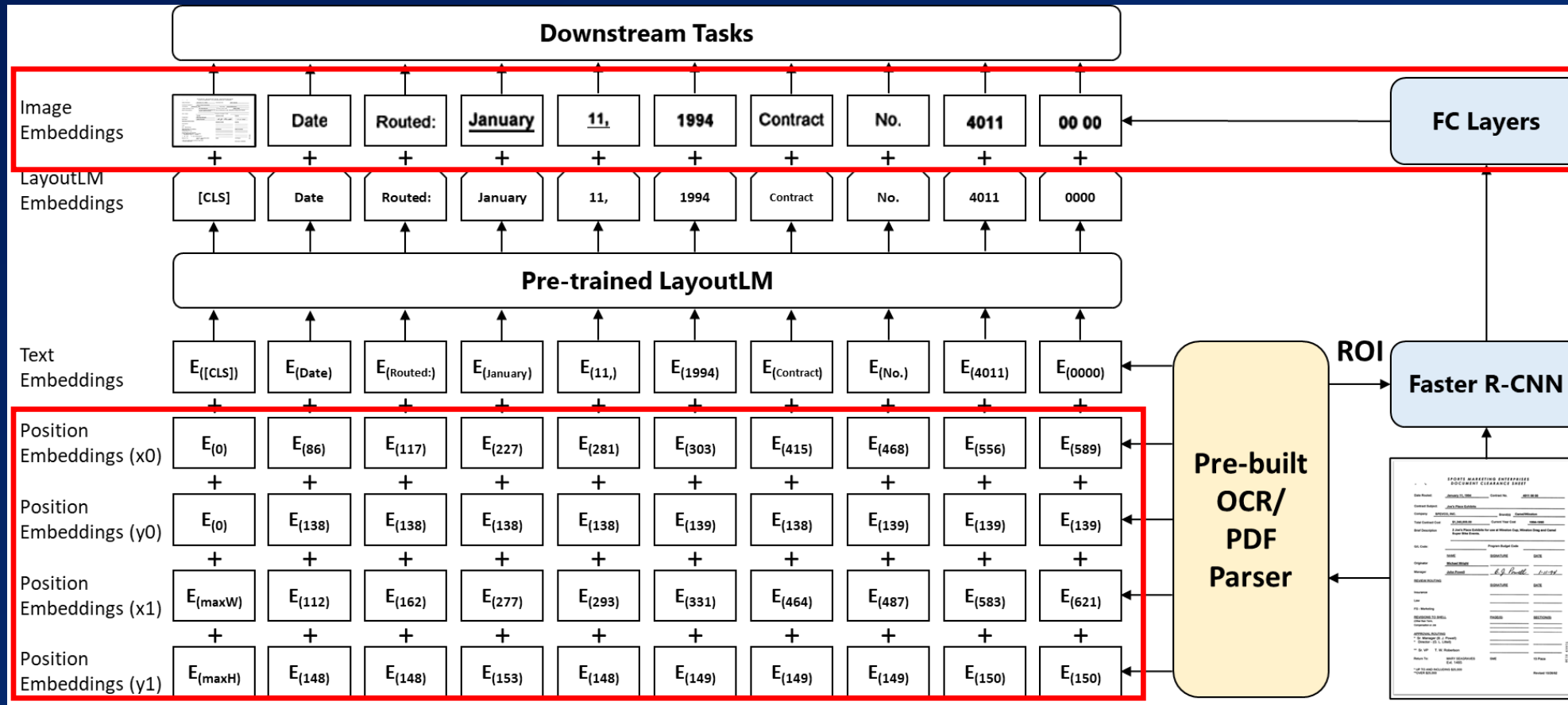
Customer ID: 0000943

Quantity	Item Code/Description	Unit Price	Extension
1.00	TENT TOP 10X10 CITRADELIC LIME	655.00	655.00
1.00	TENT TOP 10X10 CITRADELIC TANGERINE	655.00	655.00
2.00	TENT TOP 10X10 FAT TIRE	655.00	1,310.00
4.00	TENT TOP 10X10 VOODOO RANGER	655.00	2,620.00
2.00	FLAG-TEAR STD CITRADELIC LIME	177.00	354.00
2.00	FLAG-TEAR STD CITRADELIC TANGERINE	177.00	354.00
2.00	FLAG-TEAR STD FAT TIRE	177.00	354.00
4.00	FLAG-TEAR STD VOODOO RANGER	177.00	708.00
4.00	BASE-SAND/WAT SAND/WATER BASE	39.00	156.00
1.00	FREIGHT SHIPPING/TRANSPORTATION CHARGE	141.78	141.78

Subtotal: 7,500.78
Total_tax: 407.81
Total_invoice_amount: 7,515.59

- Date
- ID
- Number
- Address
- Name
- Item
- ...

LayoutLM Architecture



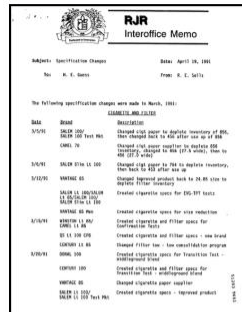
- Image Embeddings

- Position Embeddings

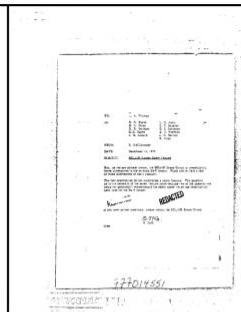
* Text embeddings initialized by BERT/UniLM

Pre-training Data

letter



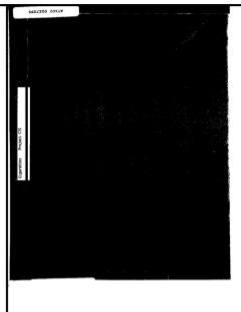
memo



email



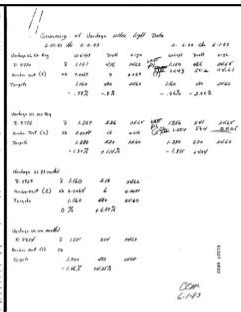
filefolder



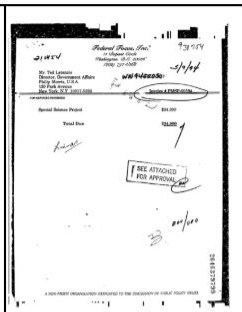
form



handwritten



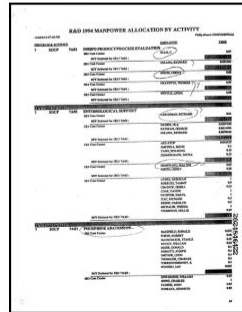
invoice



advertisement



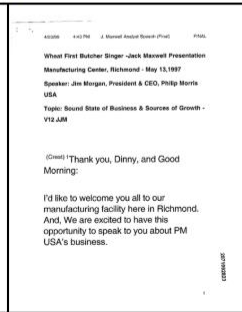
budget



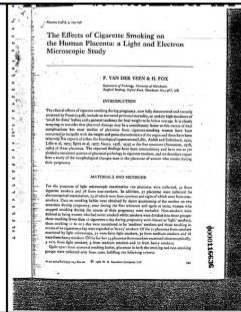
news article



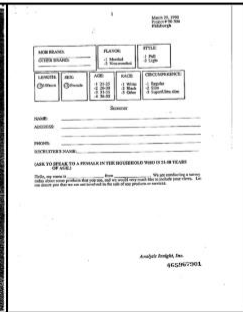
presentation



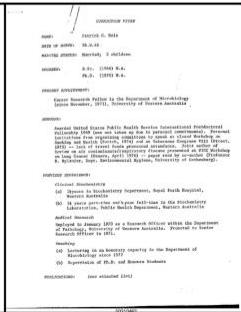
scientific publication



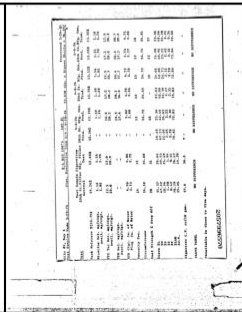
questionnaire



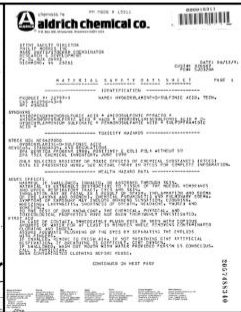
resume



scientific report



specification

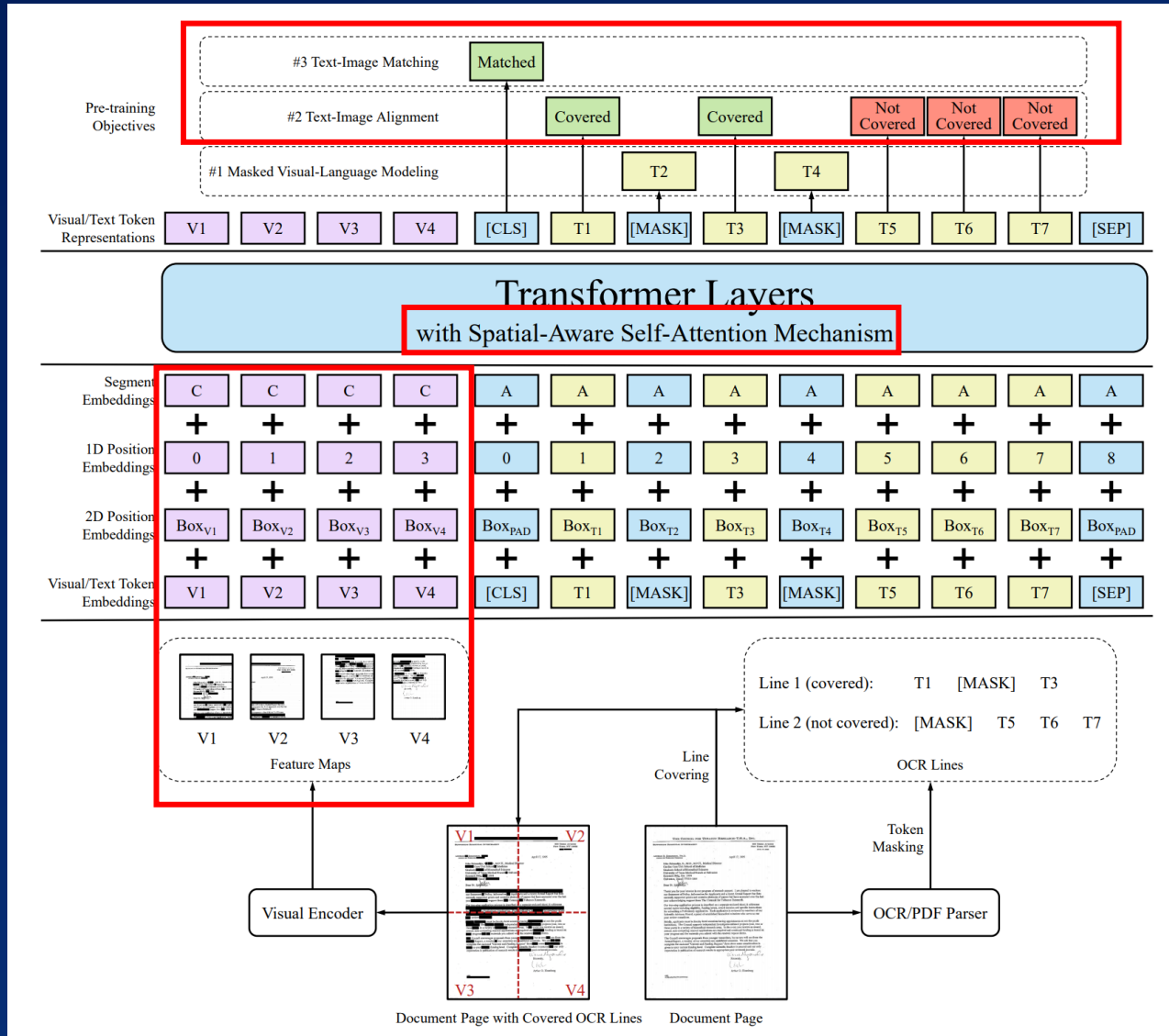


11 million scanned document images from IIT-CDIP Test Collection 1.0

<https://ir.nist.gov/cdip/>

LayoutLMv2

LayoutLMv2: Multi-modal Pre-training for Visually-Rich Document Understanding, ACL 2021



- New pre-training tasks
- New self-attention mechanism
- Image features now go through transform layers

Semantic Entity Recognition

Model	FUNSD	CORD	SROIE	Kleister-NDA
BERT _{BASE}	0.6026	0.8968	0.9099	0.7790
UniLMv2 _{BASE}	0.6648	0.9092	0.9459	0.7950
BERT _{LARGE}	0.6563	0.9025	0.9200	0.7910
UniLMv2 _{LARGE}	0.7072	0.9205	0.9488	0.8180
LayoutLM _{BASE}	0.7866	0.9472	0.9438	0.8270
LayoutLM _{LARGE}	0.7895	0.9493	0.9524	0.8340
LayoutLMv2 _{BASE}	0.8276	0.9495	0.9625	0.8330
LayoutLMv2 _{LARGE}	0.8420	0.9601	0.9781	0.8520
BROS (Hong et al., 2021)	0.8121	0.9536	0.9548	–
SPADE (Hwang et al., 2020)	–	0.9150	–	–
PICK (Yu et al., 2020)	–	–	0.9612	–
TRIE (Zhang et al., 2020)	–	–	0.9618	–
Top-1 on SROIE Leaderboard (until 2020-12-24)	–	–	0.9767	–
RoBERTa _{BASE} in (Graliński et al., 2020)	–	–	–	0.7930

Form Understanding (**FUNSD**)

<https://guillaumejaume.github.io/FUNSD/>

Receipt Understanding (**SROIE, CORD**)

<https://rrc.cvc.uab.es/?ch=13>

<https://github.com/clovaai/cord>

Document Information Extraction (**Kleister-NDA**)

<https://github.com/applicaai/kleister-nda>

Document Image Classification

Model	Accuracy	#Parameters
BERT _{BASE}	89.81%	110M
UniLMv2 _{BASE}	90.06%	125M
BERT _{LARGE}	89.92%	340M
UniLMv2 _{LARGE}	90.20%	355M
LayoutLM _{BASE} (w/ image)	94.42%	160M
LayoutLM _{LARGE} (w/ image)	94.43%	390M
LayoutLMv2 _{BASE}	95.25%	200M
LayoutLMv2 _{LARGE}	95.64%	426M
VGG-16 (Afzal et al., 2017)	90.97%	-
Single model (Das et al., 2018)	91.11%	-
Ensemble (Das et al., 2018)	92.21%	-
InceptionResNetV2 ⁶ (Szegedy et al., 2016)	92.63%	-
LadderNet (Sarkhel & Nandi, 2019)	92.77%	-
Single model (Dauphinee et al., 2019)	93.03%	-
Ensemble (Dauphinee et al., 2019)	93.07%	-

Document Image Classification (*RVL-CDIP*)

<https://www.cs.cmu.edu/~aharley/rvl-cdip/>

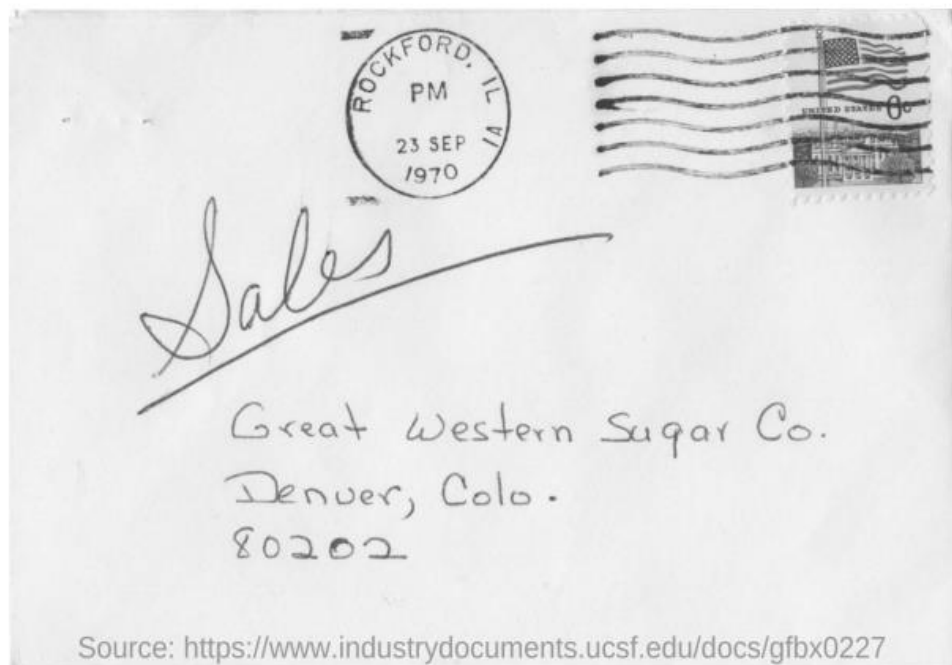
Document VQA

Model	Fine-tuning set	ANLS	#Parameters
BERT _{BASE}	train	0.6354	110M
UniLMv2 _{BASE}	train	0.7134	125M
BERT _{LARGE}	train	0.6768	340M
UniLMv2 _{LARGE}	train	0.7709	355M
LayoutLM _{BASE}	train	0.6979	113M
LayoutLM _{LARGE}	train	0.7259	343M
LayoutLMv2 _{BASE}	train	0.7808	200M
LayoutLMv2 _{LARGE}	train	0.8348	426M
LayoutLMv2 _{LARGE}	train + dev	0.8529	426M
LayoutLMv2 _{LARGE} + QG	train + dev	0.8672	426M
Top-1 on DocVQA Leaderboard (30 models ensemble) ⁷	-	0.8506	-

Document Visual Question Answering (*DocVQA*)

<https://rrc.cvc.uab.es/?ch=17>

DocVQA Leaderboard



Q: Mention the ZIP code written?

A: 80202

Q: What date is seen on the seal at the top of the letter?

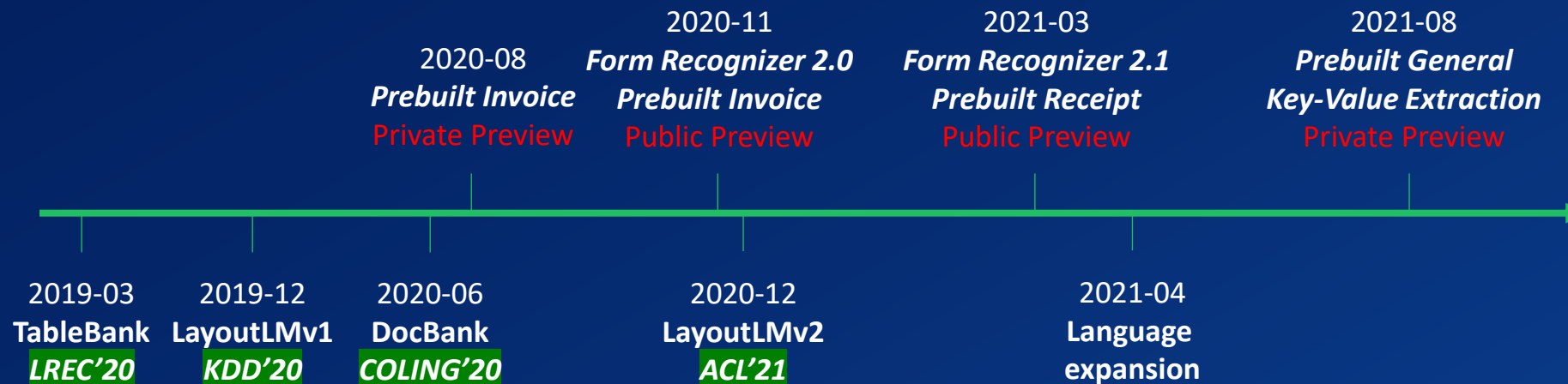
A: 23 sep 1970

Q: Which company address is mentioned on the letter?

A: Great western sugar Co.

Ranking Table ?											
<input type="checkbox"/> Description <input type="checkbox"/> Paper <input type="checkbox"/> Source Code											
Date	Method	Score	Figure/Diagram	Form	Table/List	Layout	Free_text	Image/Photo	Handwritten	Yes/No	Others
2020-06-13	Human Performance	0.9811	0.9756	0.9825	0.9780	0.9845	0.9839	0.9740	0.9717	0.9974	0.9828
2020-12-22	LayoutLM 2.0 (single model)	0.8672	0.6574	0.8953	0.8769	0.8791	0.8707	0.7287	0.6729	0.5517	0.8103
2020-08-16	Alibaba DAMO NLP	0.8506	0.6650	0.8809	0.8552	0.8733	0.8397	0.6758	0.7691	0.5492	0.7526
2020-05-16	PingAn-OneConnect-Gammlab-DQA	0.8484	0.6059	0.9021	0.8463	0.8730	0.8337	0.5812	0.7692	0.5172	0.7289
2020-05-14	Structural LM-v2	0.7674	0.4931	0.8381	0.7621	0.7924	0.7596	0.4756	0.6282	0.5517	0.6549
2020-05-15	QA_Base_MRC_2	0.7415	0.4854	0.8015	0.6738	0.7943	0.8136	0.5740	0.5831	0.5287	0.7161
2020-05-15	QA_Base_MRC_1	0.7407	0.4890	0.7984	0.6675	0.7936	0.8131	0.5854	0.6099	0.4943	0.7384
2020-05-15	QA_Base_MRC_4	0.7348	0.4735	0.8040	0.6647	0.7838	0.8043	0.5618	0.5810	0.4598	0.7332
2020-05-15	QA_Base_MRC_3	0.7322	0.4852	0.7958	0.6562	0.7842	0.8044	0.5679	0.5730	0.4511	0.7171
2020-05-15	QA_Base_MRC_5	0.7274	0.4858	0.7877	0.6550	0.7754	0.8047	0.5405	0.5619	0.4598	0.7084
2020-05-16	HyperDQA_V4	0.6893	0.3874	0.7792	0.6309	0.7478	0.7187	0.4867	0.5630	0.4138	0.5685
2020-05-16	HyperDQA_V3	0.6769	0.3876	0.7774	0.6167	0.7332	0.6961	0.4296	0.5373	0.4138	0.5650
2020-05-16	HyperDQA_V2	0.6734	0.3818	0.7666	0.6110	0.7332	0.6867	0.4834	0.5560	0.3793	0.5902
2020-05-09	HyperDQA_V1	0.6717	0.4013	0.7693	0.6197	0.7167	0.6922	0.3598	0.5596	0.4138	0.5504
2020-05-09	bert fulldata finetuned	0.5900	0.4169	0.6870	0.4269	0.6710	0.7315	0.5124	0.4900	0.4483	0.5907
2020-05-01	bert finetuned	0.5872	0.2986	0.7011	0.4849	0.6359	0.6933	0.4622	0.4751	0.4483	0.4895
2020-04-30	HyperDQA_V0	0.5715	0.3131	0.6780	0.4732	0.6630	0.5716	0.3623	0.4351	0.3793	0.4941
2020-04-27	bert	0.4557	0.2233	0.5259	0.2633	0.5113	0.7775	0.4859	0.3565	0.0345	0.5778
2020-05-16	UGLIFT v0.1 (Clova OCR)	0.4417	0.1766	0.5600	0.3178	0.5340	0.4520	0.2253	0.3573	0.4483	0.3356
2020-05-14	Plain BERT QA	0.3524	0.1687	0.4489	0.2029	0.4321	0.4812	0.3517	0.3096	0.0345	0.3747
2020-05-16	Clova OCR V0	0.3489	0.0977	0.4855	0.2670	0.3811	0.3958	0.2489	0.2875	0.0345	0.3062
2020-05-01	HDNet	0.3401	0.2040	0.4688	0.2181	0.4710	0.1916	0.2488	0.2736	0.1379	0.2458
2020-05-16	CLOVA OCR	0.3296	0.1246	0.4612	0.2455	0.3622	0.3746	0.1692	0.2736	0.0690	0.3205
2020-04-29	docVQAQV_V0.1	0.3016	0.2010	0.3898	0.3810	0.2933	0.0664	0.1842	0.2736	0.1586	0.1695
2020-04-26	docVQAQV_V0	0.2342	0.1646	0.3133	0.2623	0.2483	0.0549	0.2277	0.1856	0.1034	0.1635
2020-06-16	Test Submission	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

LayoutLM for Azure Form Recognizer



[LayoutLM: Pre-training of Text and Layout for Document Image Understanding](#), KDD'20
[LayoutLMv2: Multi-modal Pre-training for Visually-rich Document Understanding](#), ACL'21

[TableBank: A Benchmark Dataset for Table Detection and Recognition](#), LREC'20
[DocBank: A Benchmark Dataset for Document Layout Analysis](#), COLING'20

Invoice Demo

Invoice Demo

Concluding Remarks

“Universal OCR” is within our reach

- Representative training data
- Scalable computing platform and training tools
- Universal text detection and mixed language/style text recognition

Document understanding – a vision/language cross-field problem

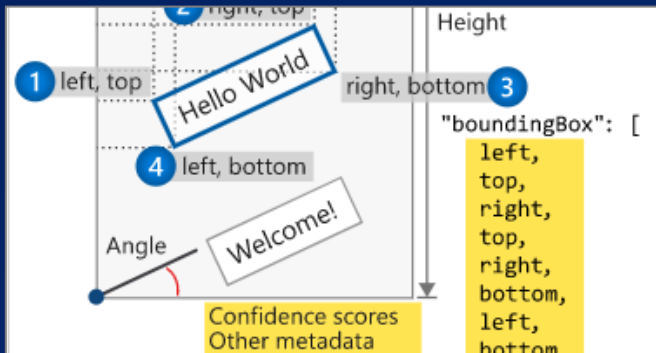
- Joint visual/language pre-training is a powerful idea
- Broadly applicable to many document understanding tasks

More researches on following topics for robotic process automation (RPA)

- Page object (especially table) detection
- Table structure recognition
- Customization

Thank you!

OCR



<https://docs.microsoft.com/en-us/azure/cognitive-services/computer-vision/overview-ocr>

Form Recognizer



<https://docs.microsoft.com/en-us/azure/cognitive-services/form-recognizer/overview>

Contact us



formrecog_contact@microsoft.com