



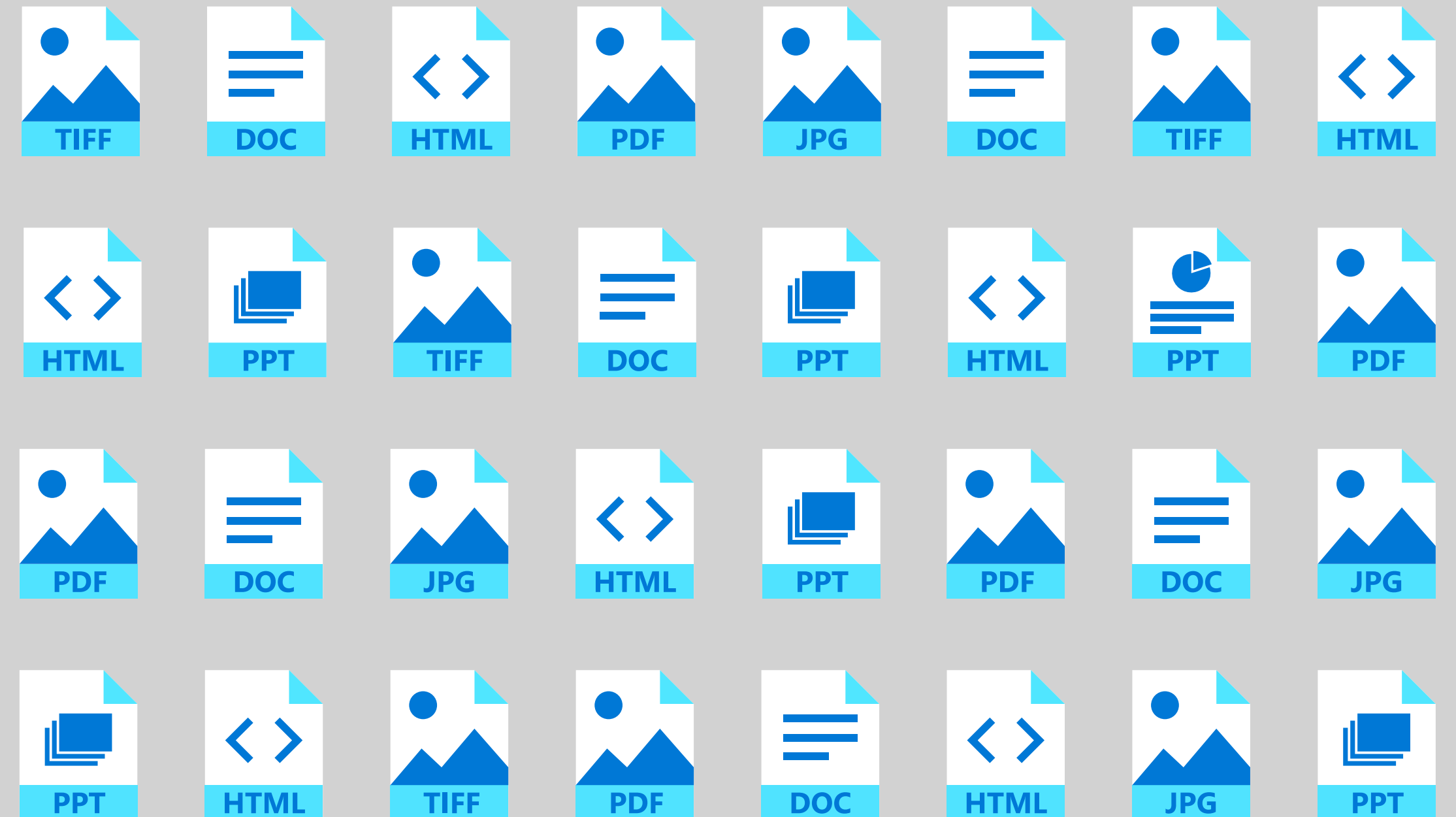
Document AI: Benchmarks, Models and Applications

Lei Cui
Microsoft Research Asia

DIL workshop, ICDAR 2021
2021-09-06

Document AI

- Analyze forms and documents
- Create intelligent search indexes
- Automate business workflows



Uncover latent insights from all your content

Document AI in Real World

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: **5/3/79** NO. **A4**

INVESTIGATOR(S): **H. S. Tong & M. S. Forte** NOTEBOOK PAGE: **BI014-23**

SIGNATURE(S): *H.S. Tong* *M.S. Forte (by D. Poole)*

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

AVERAGE WEIGHT/RANGE (GM): _____ SOURCE: **Camm 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	2592	0/6
4	10	3732	3/6
5	10	4479	6/6

REFERENCE FOR CALCULATION: **Litchfield, J. T. and Wilcoxin, 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)

Form

Scanned documents (.jpg, .png, ...)

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

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

SALE 6291458

Visa Card #XXXXXXXXXXXX8698
Magnetic card present: SICKAFOOSE/DANNY
Card Entry Method: S

Approval: 04702B

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

Receipt

UBS Global Research 31 January 2018

First Read
Microsoft Corp.
Azure Acceleration Impresses in Solid Q2

Equities
Americas Software Buy

12-month rating Buy

12m price target US\$110.00
Prior: US\$105.00
Price US\$95.01

Trading data and key metrics
52-wk range US\$95.01-63.17
Market cap. US\$734bn
Shares o/s 7.725m (COM)
Free float 90%
Avg. daily volume ('000) 23,992
Avg. daily value (m) US\$2,078.1
Common s/h equity (06/18E) US\$105bn
P/BV (06/18E) 7.0x
Net debt / EBITDA (06/18E) NM

Revenue Beat and Tax Benefits Drive Estimates Higher
Gross and operating margin upside in Q2 drives margin expectations higher for the year, with the company now expecting flat GM YoY and a slight increase in OM for FY18. Our FY18 rev. est. moves from \$10.6.6B to \$10.7.2B while FY19 goes from \$11.6.0B to \$11.7.3B. 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 SaaS multiple, which moves from 6.2x EV/CY18 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\$m)	06/15	06/16	06/17	06/18E	06/19E	06/20E	06/21E	06/22E
Revenues	93,580	91,154	96,571	107,160	117,341	129,386	142,726	157,148
EBIT (US\$m)	28,172	27,188	29,331	32,724	38,812	45,469	52,923	61,117
Net earnings (US\$m)	21,627	21,434	25,732	28,301	31,650	36,943	42,868	49,383
EPS (US\$, diluted)	2.62	2.68	3.29	3.62	4.02	4.67	5.43	6.25
DPS (US\$)	1.21	1.44	1.53	1.62	1.77	1.80	1.80	1.80
Net (debt) / cash	61,234	59,779	46,787	81,734	106,888	138,850	178,752	226,527

Profitability/valuation
EBIT margin % 30.1 29.8 30.4 30.5 33.1 35.1 37.1 38.9
ROIC (EBIT) % 29.9 27.4 122.8 120.9 173.4 157.4 148.3 141.6
EV/EBITDA (core) x 8.7 9.8 11.4 15.5 12.7 10.5 8.6 6.9
P/E (US\$, diluted) x 17.2 18.8 19.0 26.2 23.6 20.3 17.5 15.2
Equity FCF (US\$) yield % 6.2 6.1 6.4 5.9 4.3 5.1 6.1 7.0
Net dividend yield % 2.7 2.9 2.5 1.7 1.9 1.9 1.9 1.9

Source: Company accounts, Thomson Reuters, UBS estimates. US\$ adjusted EPS is stated before goodwill-related charges and other adjustments for abnormal and economic items at the analyst's judgment. Valuations based on an average share price that year. (E) based on a share price of 109.05 on 31 Jan. 2018 18:48 EST

Report

Digital-born documents (.pdf, .docx, ...)

Access Information Protected. 10445 48th Avenue Denver, CO 80228

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

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

Service Billing Period: 1/31/2017
Date: 1/31/2017
Invoice #: 1861619
Customer #: GDP00286

Total Amount Due: **\$546.69**
By: 3/2/2017
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	FEE
Storage						
Storage Period: 02/01/2017 - 02/28/2017						
4	Legal Bankers Box		10.00	0.5040	N	5.04
468	Letter Bankers Box		936.00	0.5040	N	471.75
85	Letter Legal Box		85.00	0.5400	N	45.90
			TOTAL FOR Storage			522.69
			TAX			0.00
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
		FileBRIDGE Records + AccessMETRICS	1.00	0.0000	N	0.00
			TOTAL FOR Service			24.00
			TAX			0.00
Transportation						
		Shred Rotation Transportation - Scheduled trip	2.00	0.0000	N	0.00
			TOTAL FOR Transportation			0.00
			TAX			0.00
			SUB-TOTAL			546.69
			TAX			0.00
			INVOICE TOTAL			\$546.69

PLEASE NOTE: To the extent you do not have a currently effective written contract for services with an Access or Retrieval company, by paying this invoice, you agree that the terms and conditions found on <http://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 Retrieval 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 (BASA) with an Access or Retrieval 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 Retrieval 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 Retrieval are not binding.

Invoice

Layout invariance (key-value, tabular, etc.) among visually-rich documents

Document AI Tasks

Key	Value
TO	Lorillard Corporation
ADDRESS	666 Fifth Avenue
CITY	New York
...	...

Form Understanding

Key	Value
Total	
Company	
Address	
Date	



Key	Value
Total	4.95
Company	Starbucks Store
Address	11302 Euclid Avenue Cleveland, OH
Date	12/07/2014

Receipt Understanding

Category: Form

Document Image Classification

Document AI Products

Applications

Key-value Extraction
Document Classification
Document VQA

Downstream Tasks

Table Detection
Page Object Detection
Reading Order Detection

Benchmarks

TableBank
(LREC'20)

DocBank
(COLING'20)

ReadingBank
(EMNLP'21)

XFUND
Benchmark

Foundation Models

LayoutLM/LayoutLMv2/LayoutXLM
(KDD'20, ACL'21)

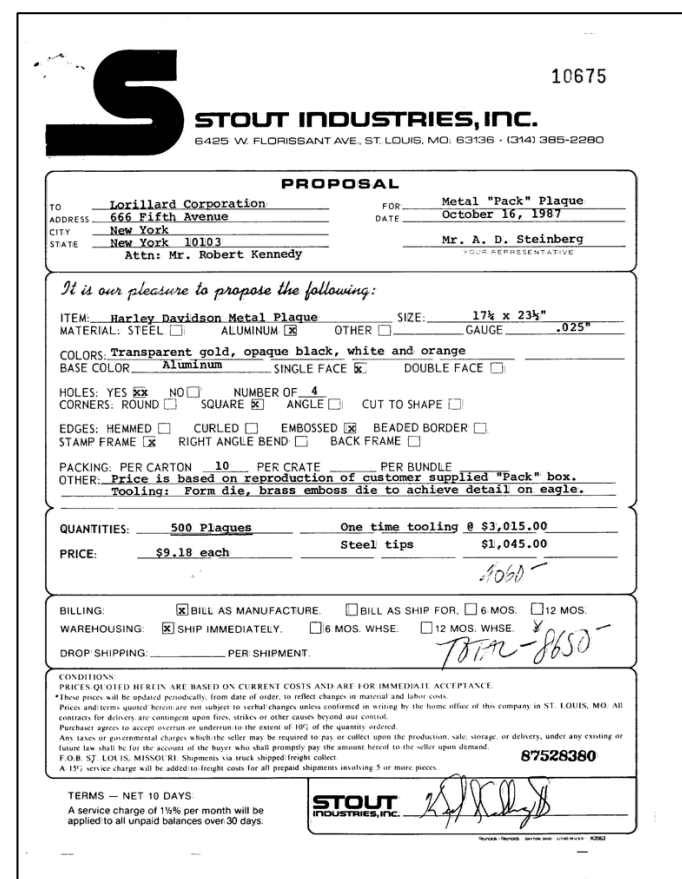
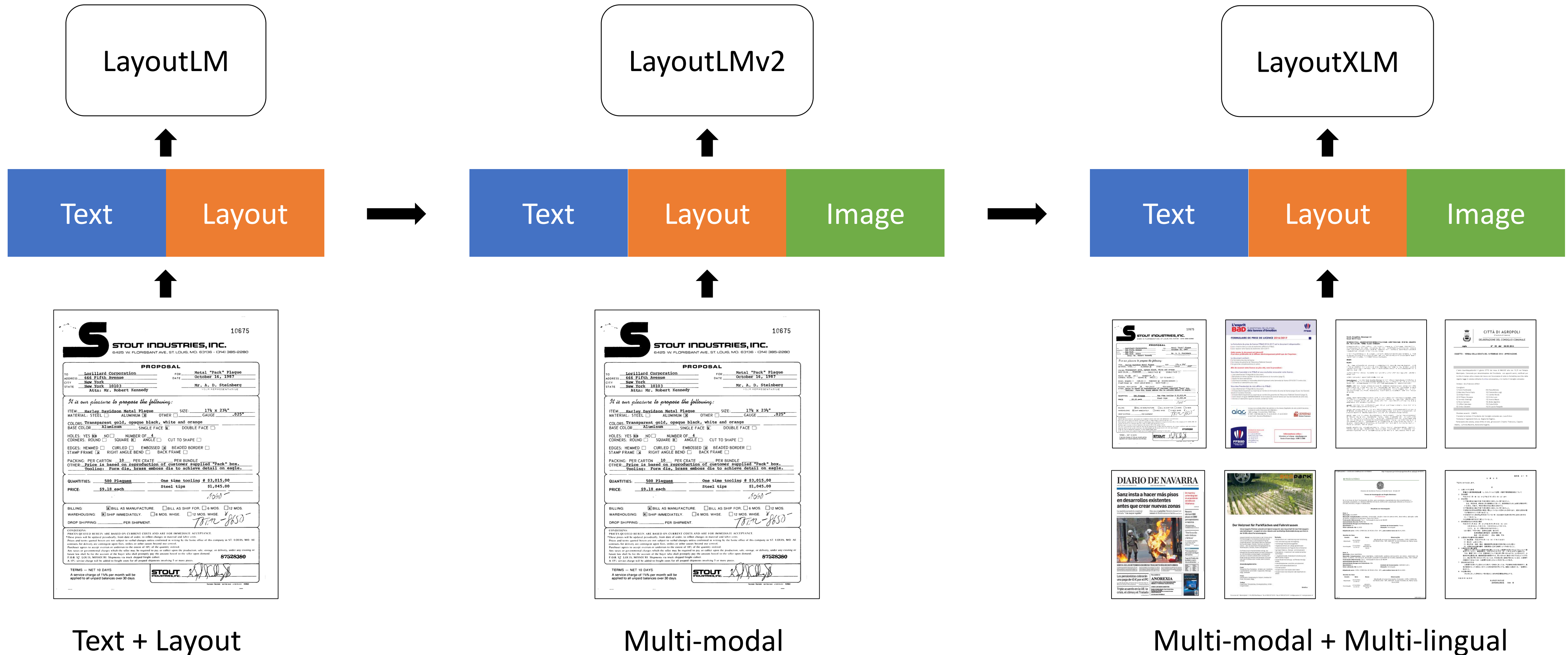
<https://aka.ms/document-ai/>



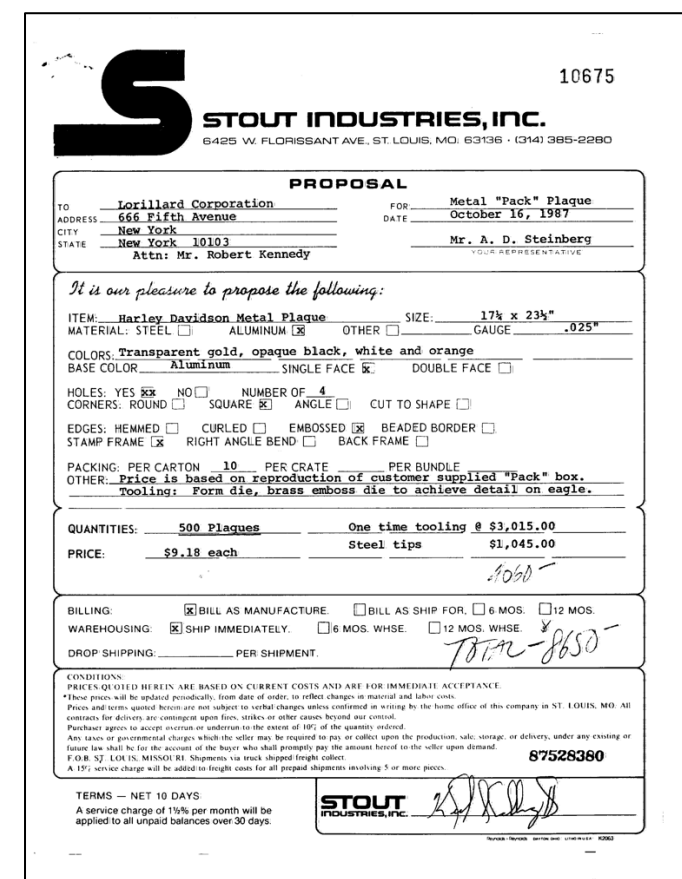
LayoutLM Model Family



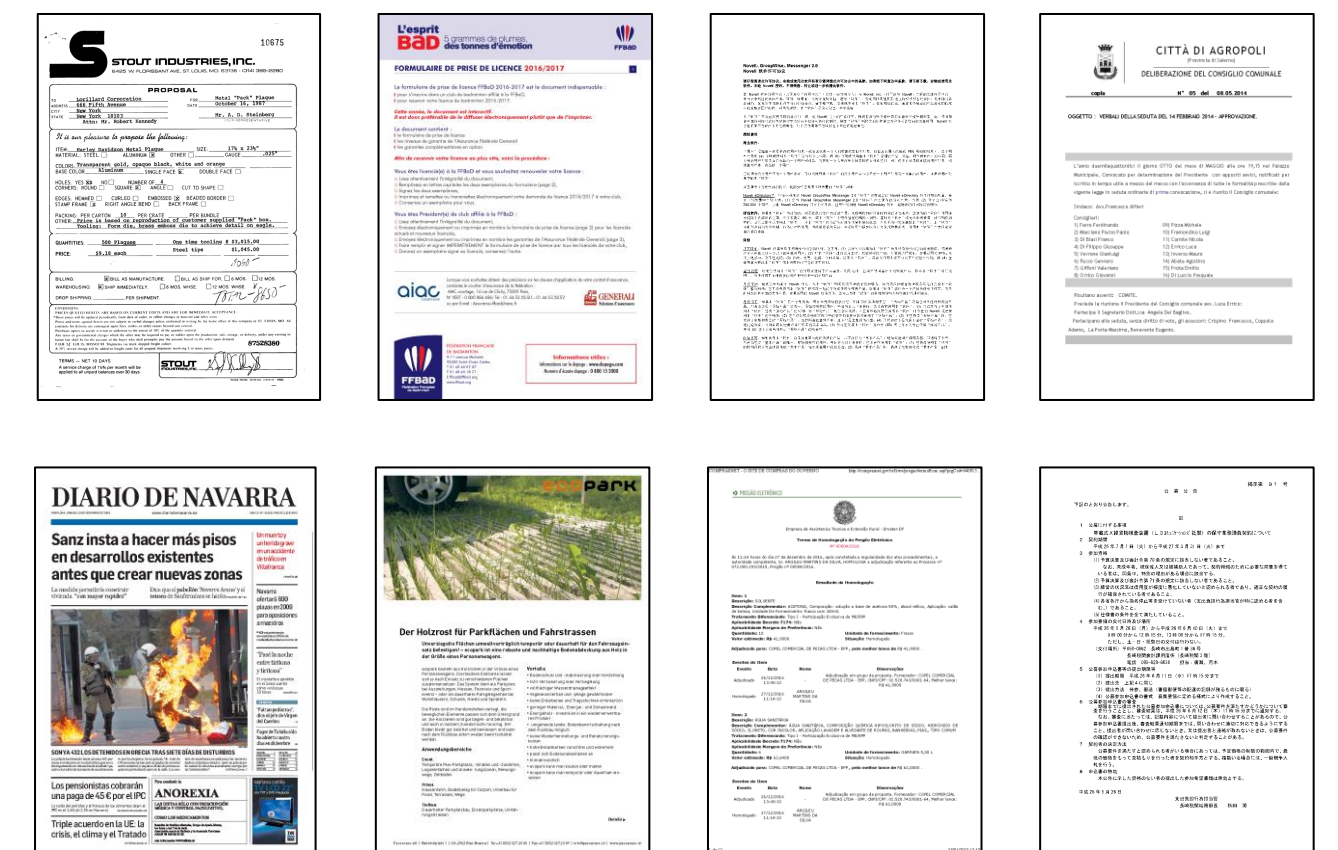
LayoutLM -> LayoutLMv2 -> LayoutXLM



Text + Layout

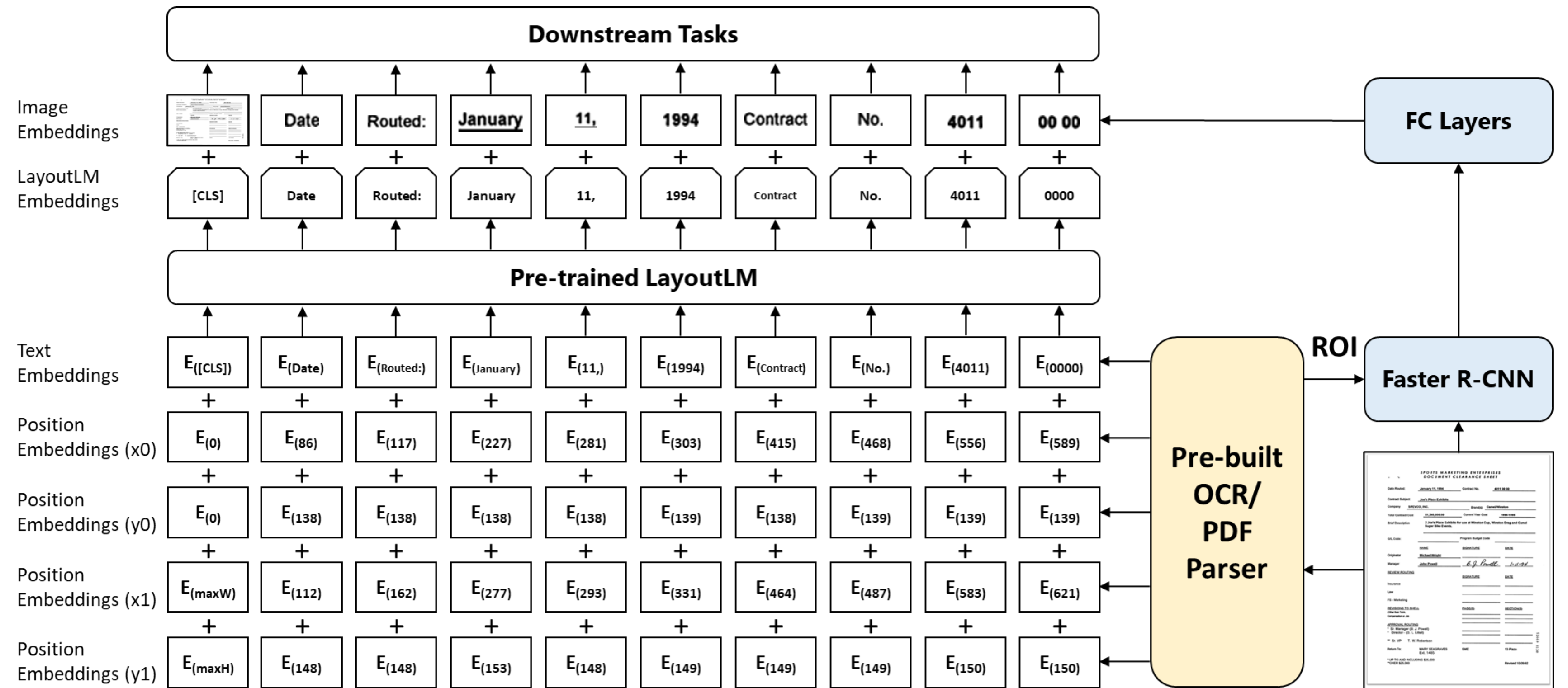


Multi-modal



Multi-modal + Multi-lingual

LayoutLM



* Text embeddings initialized by BERT/UniLM

**SPORTS MARKETING ENTERPRISES
DOCUMENT CLEARANCE SHEET**

Date Routed: January 11, 1994 Contract No. 4011 00 00

Contract Subject: Joe's Place Exhibits

Company SPEVCO, INC. Brand(s) Camel/Winston

Total Contract Cost \$1,340,000.00 Current Year Cost 1994-1995

Brief Description 2 Joe's Place Exhibits for use at Winston Cup, Winston Drag and Camel Super Bike Events.

G/L Code: _____ Program Budget Code _____

	<u>NAME</u>	<u>SIGNATURE</u>	<u>DATE</u>
Originator	<u>Michael Wright</u>	_____	_____
Manager	<u>John Powell</u>	<u>B. J. Powell</u>	<u>1-11-94</u>

REVIEW ROUTING

	<u>SIGNATURE</u>	<u>DATE</u>
Insurance	_____	_____
Law	_____	_____
FS - Marketing	_____	_____

REVISIONS TO SHELL
(Other than Term, Compensation or Job)

	<u>PAGE(S)</u>	<u>SECTION(S)</u>
_____	_____	_____
_____	_____	_____

APPROVAL ROUTING

- * Sr. Manager (B. J. Powell)
- * Director - (G. L. Littell)

** Sr. VP T. W. Robertson

Return To: MARY SEAGRAVES SME 13 Plaza
Ext. 1485

* UP TO AND INCLUDING \$25,000
**OVER \$25,000

Revised 10/26/92

51669 8130

Date Routed: January 11, 1994 Contract No. 4011 00 00

↓ OCR/PDF Parser

Image	Token	Bounding Box (x0,y0,x1,y1)
Date	Date	86 138 112 148
Routed:	Routed:	117 138 162 148
<u>January</u>	January	227 138 277 153
11	11,	281 138 293 148
1994	1994	303 139 331 149
Contract	Contract	415 138 464 149
No.	No.	468 139 487 149
4011	4011	556 139 583 150
00 00	0000	589 139 621 150

↓

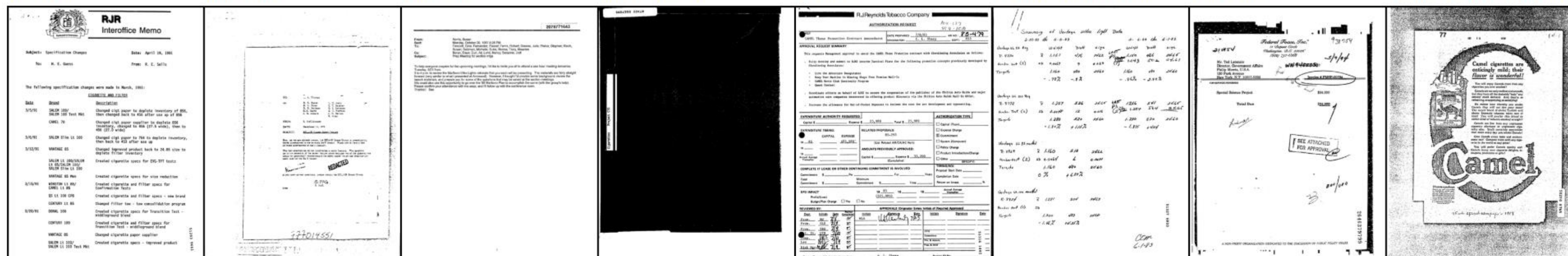
Input

Tok 0	Tok 1	Tok 2	Tok 3	Tok 4	Tok 5	Tok 6	Tok 7
-------	-------	-------	-------	-------	-------	-------	-------

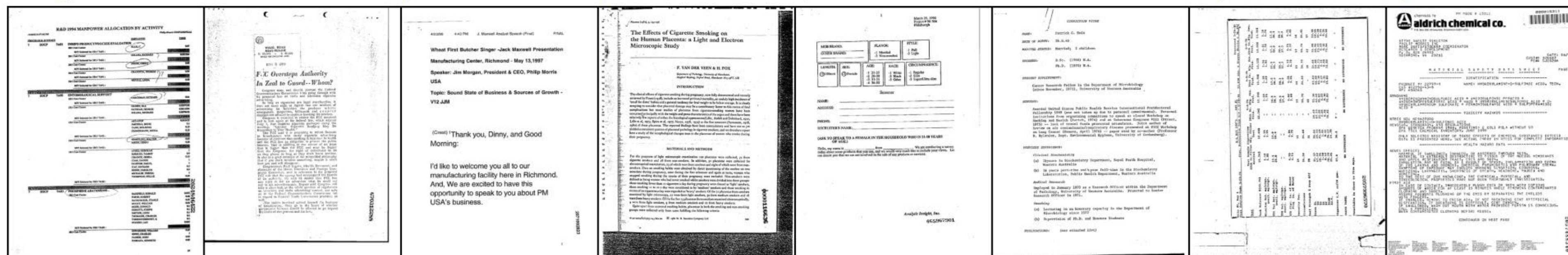
Token Embeddings	$E_{(Tok 0)}$	$E_{(Tok 1)}$	$E_{(Tok 2)}$	$E_{(Tok 3)}$	$E_{(Tok 4)}$	$E_{(Tok 5)}$	$E_{(Tok 6)}$	$E_{(Tok 7)}$
	+	+	+	+	+	+	+	+
Position Embeddings (x0)	$E_{(Tok 0)}$	$E_{(Tok 1)}$	$E_{(Tok 2)}$	$E_{(Tok 3)}$	$E_{(Tok 4)}$	$E_{(Tok 5)}$	$E_{(Tok 6)}$	$E_{(Tok 7)}$
	+	+	+	+	+	+	+	+
Position Embeddings (y0)	$E_{(Tok 0)}$	$E_{(Tok 1)}$	$E_{(Tok 2)}$	$E_{(Tok 3)}$	$E_{(Tok 4)}$	$E_{(Tok 5)}$	$E_{(Tok 6)}$	$E_{(Tok 7)}$
	+	+	+	+	+	+	+	+
Position Embeddings (x1)	$E_{(Tok 0)}$	$E_{(Tok 1)}$	$E_{(Tok 2)}$	$E_{(Tok 3)}$	$E_{(Tok 4)}$	$E_{(Tok 5)}$	$E_{(Tok 6)}$	$E_{(Tok 7)}$
	+	+	+	+	+	+	+	+
Position Embeddings (y1)	$E_{(Tok 0)}$	$E_{(Tok 1)}$	$E_{(Tok 2)}$	$E_{(Tok 3)}$	$E_{(Tok 4)}$	$E_{(Tok 5)}$	$E_{(Tok 6)}$	$E_{(Tok 7)}$

Pre-training Data

letter memo email filefolder form handwritten invoice advertisement



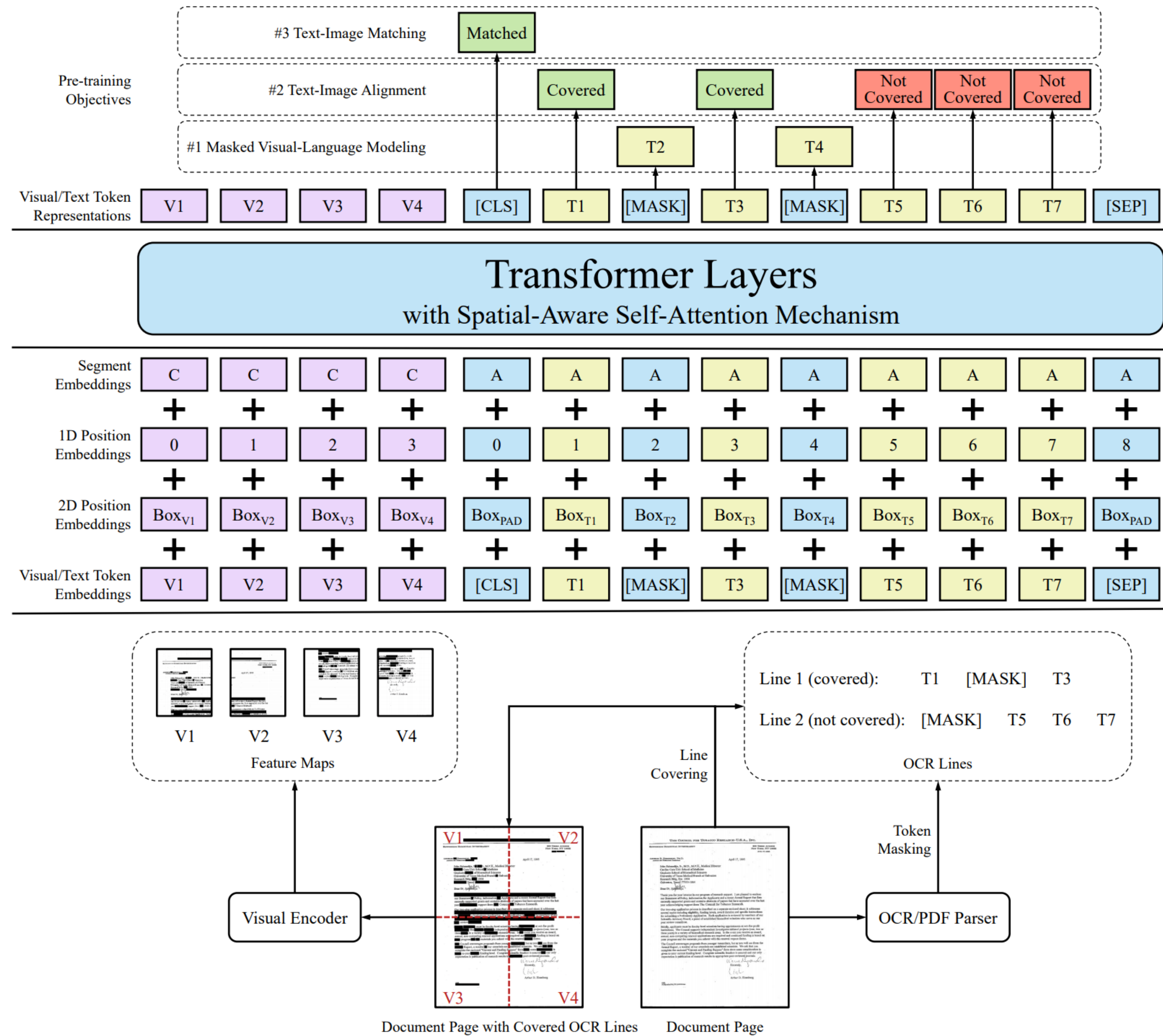
budget news article presentation scientific publication questionnaire resume scientific report specification



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

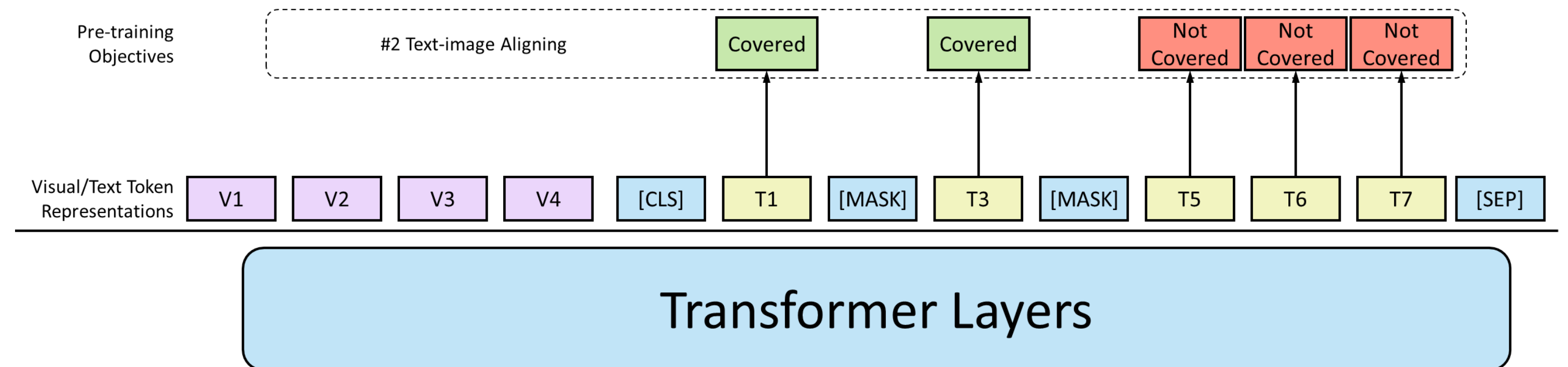
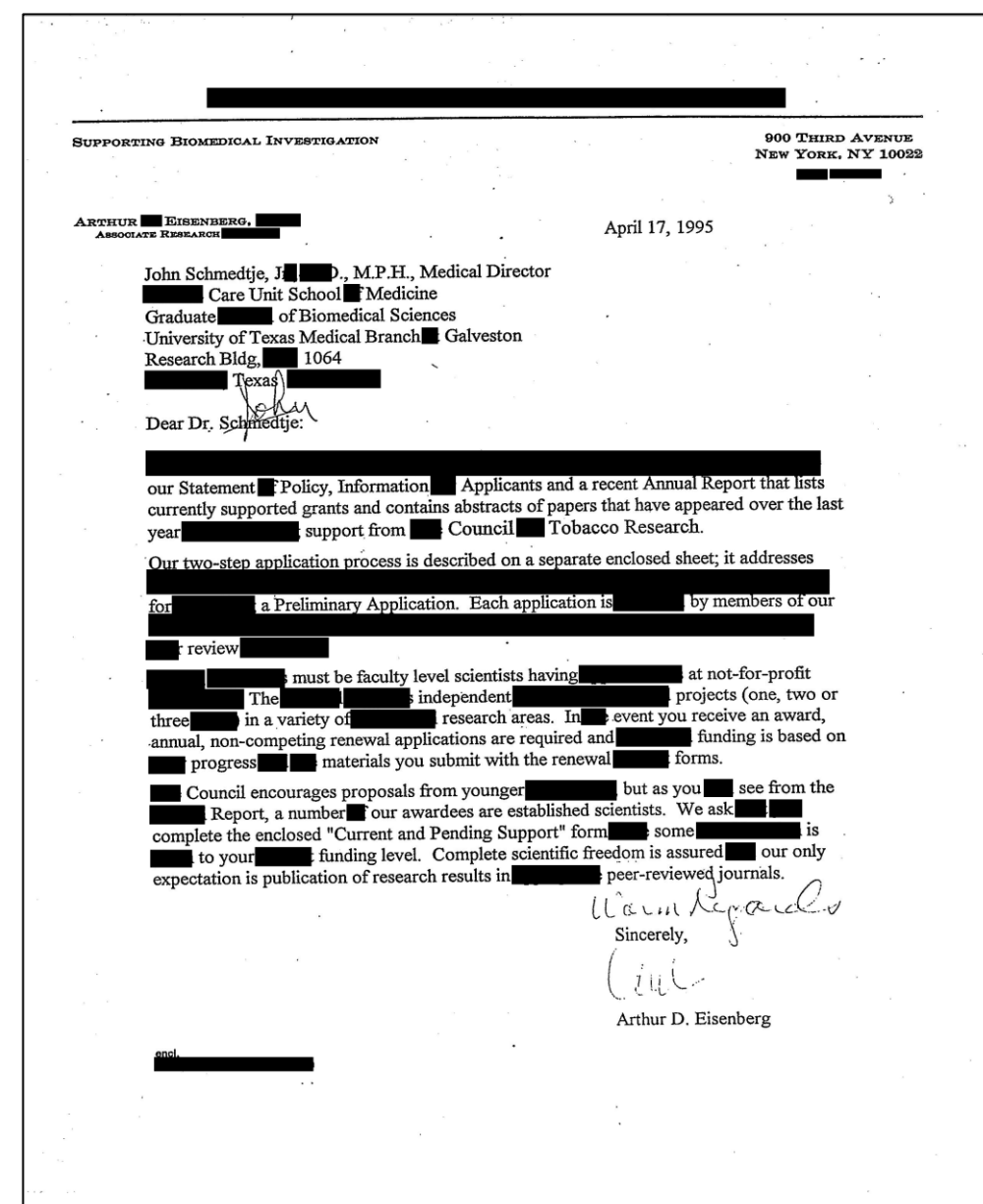
<https://ir.nist.gov/cdip/>

LayoutLMv2

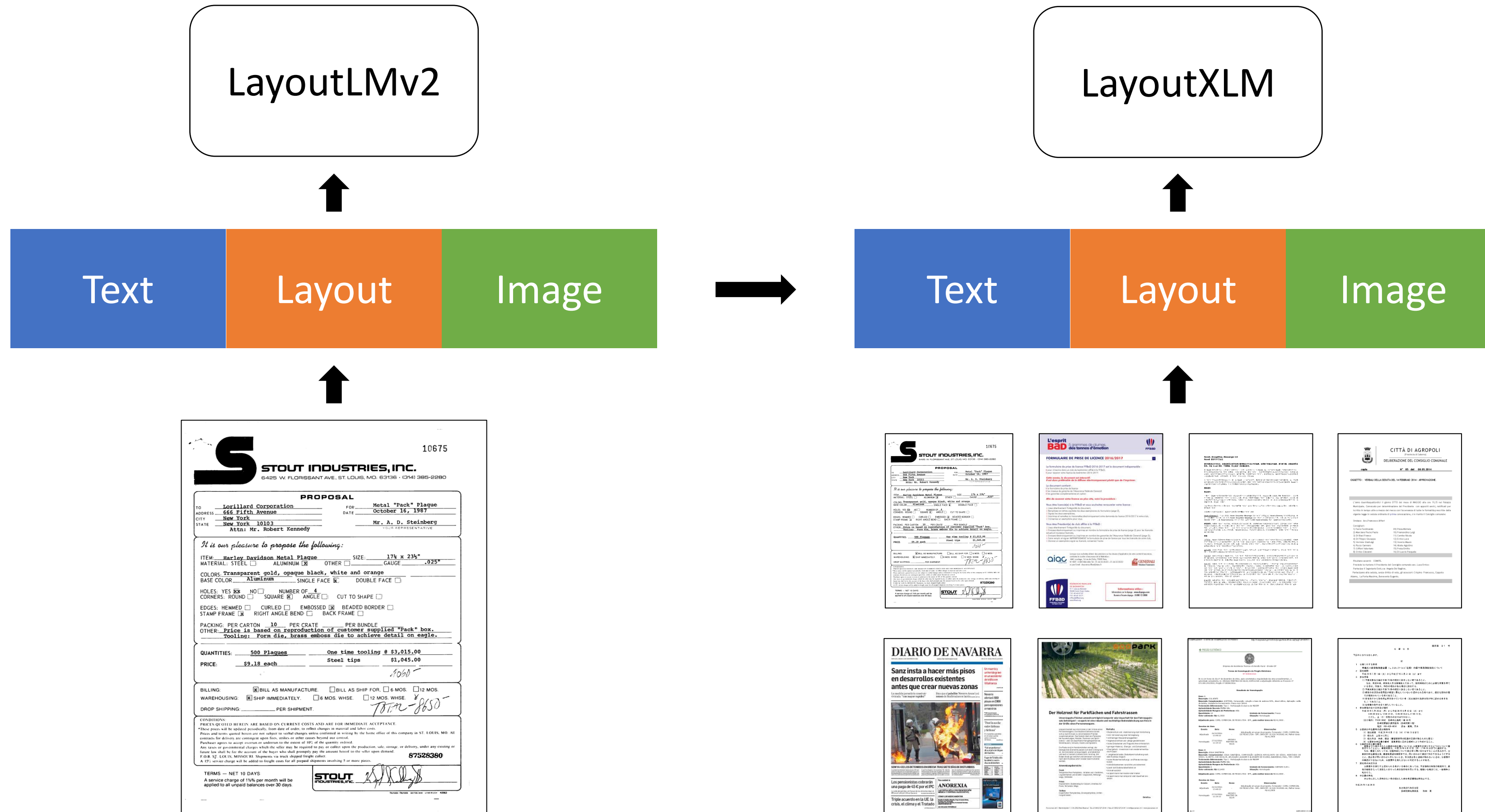


Pre-training Tasks

- Masked Visual-Language Modeling
- Text-Image Matching
- **Text-Image Alignment**



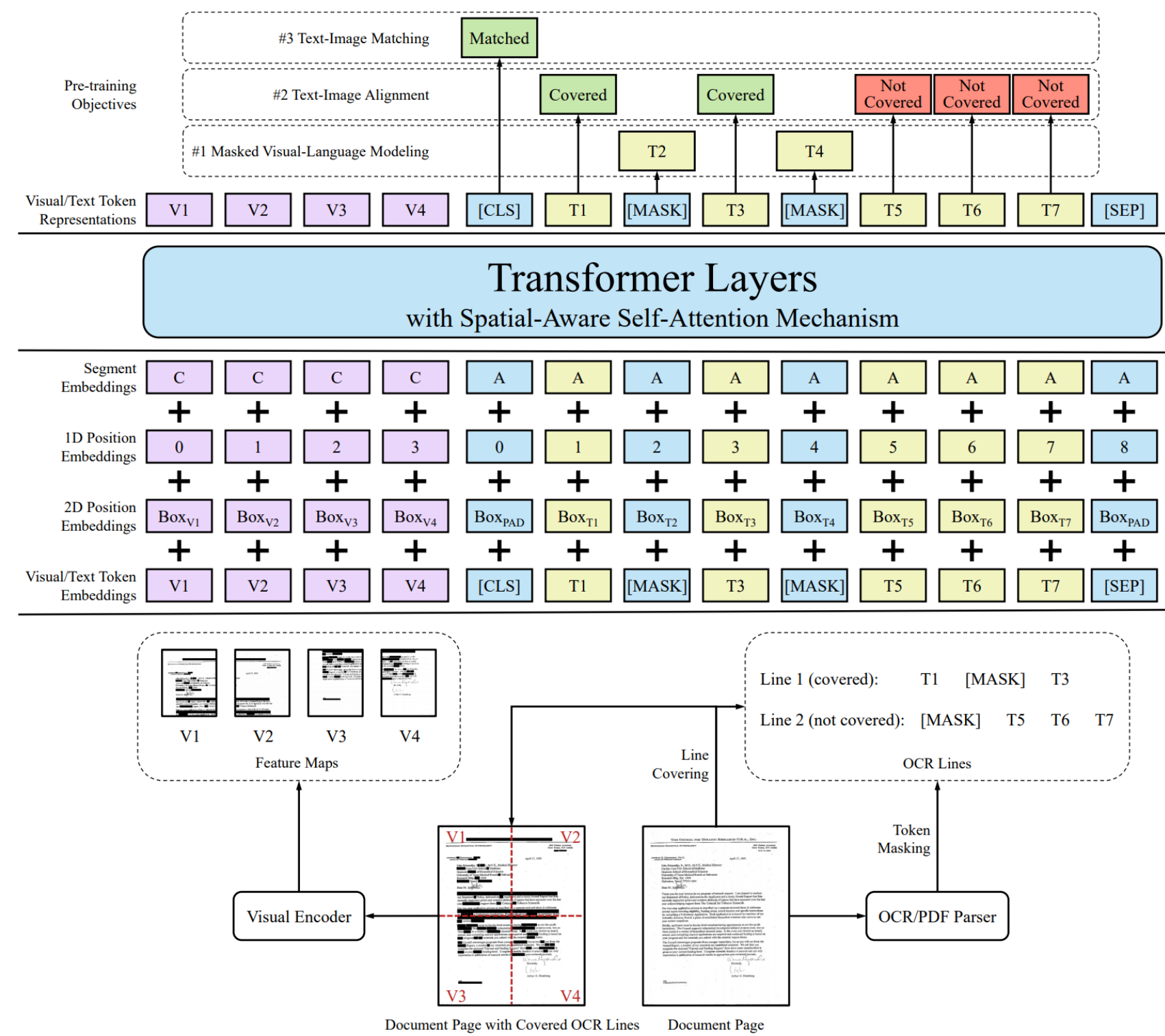
LayoutLMv2 -> LayoutXLM



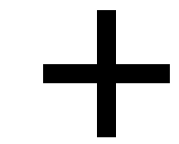
Multi-modal

Multi-modal + Multi-lingual

LayoutXLM



LayoutLMv2 + InfoXLM



(a) English

(b) Chinese

(c) Japanese

(d) Spanish

(e) French

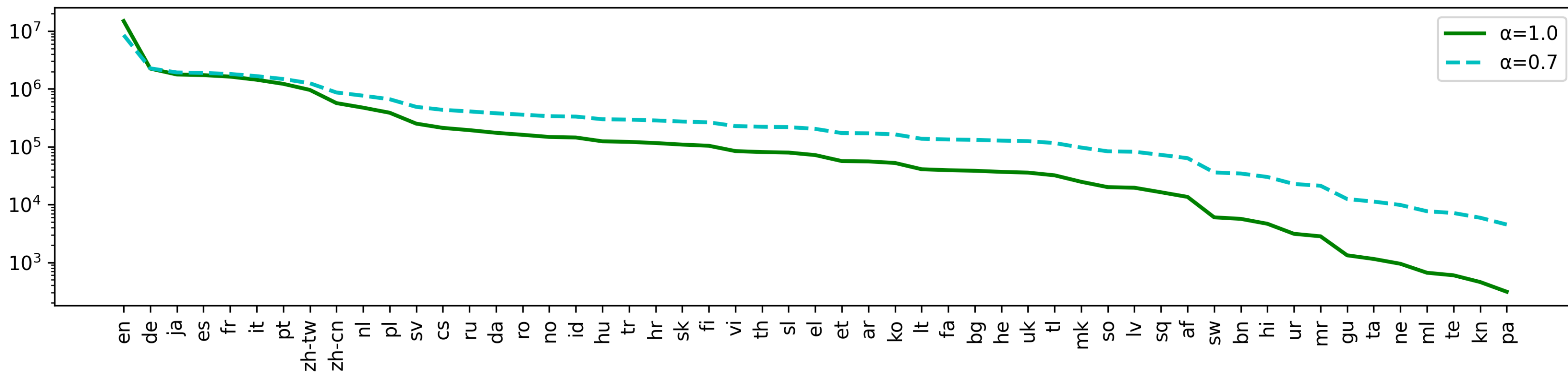
(f) Italian

(g) German

(h) Portuguese

Multi-lingual documents (50+ languages)

Language Distribution for Pre-training



Totally **30M document images with 50+ languages** used for pre-training the **LayoutXLM**



Hugging Face
@huggingface



Document parsing meets 🤗 Transformers!

#LayoutLMv2 and #LayoutXLM by @MSFTResearch are now available! 🔥

They're capable of parsing document images (like PDFs) by incorporating text, layout, and visual information, as in the @gradio demo below ↓

huggingface.co/spaces/nielsr/...

Interactive demo: LayoutLMv2

Demo for Microsoft's LayoutLMv2, a Transformer for state-of-the-art document image understanding tasks. This particular model is fine-tuned on FUNSD, a dataset of manually annotated forms. It annotates the words appearing in the image as QUESTION/ANSWER/HEADER/OTHER. To use it, simply upload an image or use the example image below and click 'Submit'. Results will show up in a few seconds. If you want to make the output bigger, right-click on it and select 'Open image in new tab'.

image

annotated image

21K views

0:06 / 0:09

Clear Submit Screenshot Flag

10:52 PM · Aug 31, 2021 · Twitter Web App

193 Retweets 18 Quote Tweets 716 Likes

microsoft/layoutlm-base-uncased
Updated Aug 11 · 448k · ❤️ 3

microsoft/layoutlmv2-base-uncased
Updated Aug 16 · 54.6k · ❤️ 5

microsoft/layoutxlm-base
Updated Aug 26 · 9.12k · ❤️ 6



Benchmark Datasets



183.102

154.178

TableBank

13°C and the natural ventilation reduced it further by 5°C. The indoor operative temperature will not rise above 26°C, which is the requirement for the building. Looking at the impact of the shading device coupled with natural ventilation, it can be assumed that this should solve most of the overheating problems, when the temperature rises over 24°C.

4.2.3 Visual control

The glare analysis was performed only for two rooms on two opposite facades, room 1 and room 6. The simulations were carried out from the point of view of a person staying in bed. Annual Daylight Glare Probability simulation checks every hour of the year and if the DGP was higher than 0.35, it was noted on the graph in grey scale. Darker colour means bigger glare issues, black colour means that the glare is intolerable and DGP is over 0.45. The figure below shows the results when there is no shading applied, also during shading season and when the person turns the head away from the window.

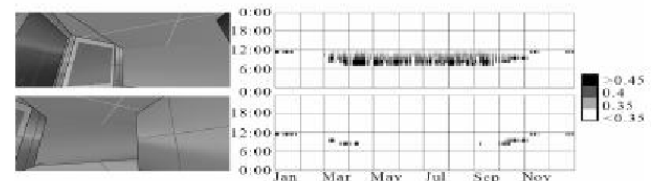


Figure 4-21 Results of annual Daylight Glare Probability for room 1 in case of normal view and view when looking away from the window

On the left side of the figure 4-21 there are approximate fields of view from the point where the head of the observer is. First one is for the person looking at the window, second is when the person looks away. On the right side of the figure there are temporal maps of Daylight Glare Probability for both views. Turning the head away solves almost all problems but the remaining issues are in the range of intolerable glare. In this case it is 69 hours of intolerable glare. Applying the shading during summer and additionally lowering it during the hours when there is direct sunlight reduces the amount of hours with intolerable glare to 13h without looking away. However lowering the shading reduces the solar gains during the heating season and increases the heating from 9.4 to 10.9 kWh/(m²·year), thus by 16%. It will also influence the amount of light in the room, which will affect the electric lighting design. The table below shows the effects of keeping the head away and applying shading on heating demand in base case for Room 1 and in the other two cases for room 6.

Table 4-4 Results of annual DGP and corresponding heating demand when using chosen external shading device as a visual control device

DGP, hours over 0.35	Room 1-orient		Room 1-base case		Room 6-orient		Room 6-base case	
	Shading	Away	Shading	Away	Shading	Away	Shading	Away
	13	69	36	94	15	183	11	202

ID	NAME	ID	PROGRAM
21	MUHO TAUFIR BIN HALEY	002039F	CYBER-CIVIL ENGINEERING
22	MUHO TAUFIR BIN HALEY	002039F	CYBER-CIVIL ENGINEERING
23	AHMAD FARIS BIN ROSMAN	002040F	CYBER-CIVIL ENGINEERING
24	MUHAMMAD ZUBIR BIN SAUDIRNOON	002043F	CYBER-CIVIL ENGINEERING
25	HUSEIN FAREED GASYPOV	002050F	CYBER-CIVIL ENGINEERING
26	SUNIL MOHAMMED GUBARI	002051F	CYBER-CIVIL ENGINEERING

SEMESTER MAY 2015: ACADEMIC WRITING

GROUP: 7

TIME/VENUE: MONDAY 9 – 11; LR 5

LECTURER: DR ENA BHATTACHARYYA

ID	NAME	ID	PROGRAM
1	SHUJAKSHY SERRAARSH	001901F	EE-4000 ELECTRICAL & ELECTRONIC
2	CHYO SHAM YAU	001902F	EE-4000 ELECTRICAL & ELECTRONIC
3	WANG CHIE YUEN	001905F	EE-4000 ELECTRICAL & ELECTRONIC
4	LOI WEN JUN	001906F	EE-4000 ELECTRICAL & ELECTRONIC
5	JACKY LING NAN CHYI	001906F	EE-4000 ELECTRICAL & ELECTRONIC
6	MUHAMMAD SAUJULAH BIN MUHO SALMAN	001907F	EE-4000 ELECTRICAL & ELECTRONIC
7	MUHAMMAD ZUBIR BIN SAUDIRNOON	001907F	EE-4000 ELECTRICAL & ELECTRONIC
8	NURAN BIN ABDUL RAHIM	001907F	EE-4000 ELECTRICAL & ELECTRONIC
9	HARYUDION ARIAN BIN HAUS	001908F	EE-4000 ELECTRICAL & ELECTRONIC
10	MUHO SYAZWAN ARIFF BIN MUHO SAUF	001908F	EE-4000 ELECTRICAL & ELECTRONIC
11	CHAN WEE SUNG	001908F	EE-4000 ELECTRICAL & ELECTRONIC
12	AMAN DANIAL B. USMO KHAN	001909F	EE-4000 ELECTRICAL & ELECTRONIC
13	WAN MUHAMMAD AYYAZ BIN WAN ZULFAIRHAZI	001909F	EE-4000 ELECTRICAL & ELECTRONIC
14	HADIRAH BT HANI	001910F	EE-4000 ELECTRICAL & ELECTRONIC
15	MUHO ATRIQ BIN MUHO MALLANA	001910F	EE-4000 ELECTRICAL & ELECTRONIC
16	CHIEWNG CHIE WEI	001910F	EE-4000 ELECTRICAL & ELECTRONIC
17	HSHAN BIN MUHAMMAD JETON ZABI	001910F	EE-4000 ELECTRICAL & ELECTRONIC
18	JAY ANIL SINGH SHEKHAR	001911F	EE-4000 ELECTRICAL & ELECTRONIC
19	AHMAD FARIS BIN ROSMAN	001912F	EE-4000 ELECTRICAL & ELECTRONIC

Table 3: Results of the component classifiers on the Airbus using five weighting schemes

Matrix	Accuracy	Precision	Recall	F-measure	Iterations	Time in seconds
Frequency	63.16	61.22	71.9	66.14	13	1
Frequency-AW	83.84	87.33	78.03	82.42	19	1
Frequency-IDF	88.94	87.61	79.09	83.14	12	1
Frequency-TN	61.29	58.09	76.43	66.4	13	1
Frequency-TFIDF	79.95	84.19	73.77	78.64	13	0
Frequency-WFIDF	83.48	87.04	78.7	82.66	12	1
Presence	79.95	79.55	69.91	74.42	17	0
Presence-AW	83.88	81.63	82.82	82.71	11	0
Presence-IDF	84.21	85.09	81.76	83.52	9	1
Presence-TN	62.00	59.87	73.5	65.99	11	1
Presence-TFIDF	82.54	85.49	78.43	81.81	16	1
Presence-WFIDF	83.54	84.33	82.42	83.37	12	0

Table 4: Results of the component classifiers on the HomeBuilders using five weighting schemes

Matrix	Accuracy	Precision	Recall	F-measure	Iterations	Time in seconds
Frequency	84.55	91.03	88.18	86.36	9	1
Frequency-AW	94.5	95.62	93.27	94.43	11	1
Frequency-IDF	94.5	95.79	93.09	94.42	8	1
Frequency-TN	74.5	71.65	81.09	76.08	7	3
Frequency-TFIDF	92.27	95.06	89.18	92.03	15	1
Frequency-WFIDF	94.45	95.62	93.18	94.38	8	1
Presence	88.50	96.4	81.09	87.67	8	2
Presence-AW	98.95	98.66	92.09	95.84	13	1
Presence-IDF	98.73	98.64	91.64	95.09	9	1
Presence-TN	82.95	86.29	78.36	82.13	7	2
Presence-TFIDF	98.86	98.73	91.82	95.24	11	1
Presence-WFIDF	98.95	98.48	92.27	95.85	7	1

learning process [23, 29]. Therefore, for enhancing the algorithm's performance, we conduct feature reduction via matching all adjectives and adverbs against SWN. Since we are interested in positive and negative classes only polar features are considered, and the reduction is done by removing neutral terms because they do not carry the clustering characteristic of reviews. When applying feature reduction on Airbnos and HomeBuilders datasets, there are slight changes which are shown in Figure 6.

Sentiment scores. In Figure 7, the sentiment scores from SWN are added to all the matrices. The polarity score has a negative impact on accuracy which was anticipated because the sentiment score is the average score of the words to which a term belongs, and the context in which a term occurs, is not considered. However, the average score is likely to correctly indicate the term polarity, that is, whether it is positive, negative or neutral. This step doubles the number of the vector space models which improves the ensemble method by promoting the groups' identification.

Experiments on multi-domain datasets. In this section, we present the results of ACACCE on different domains datasets. After applying the contextual analysis and constructing the matrices, the last step is to feed the matrices into the ensemble method, in which a document will be classified as a positive/negative instance if the majority of the com-

Table 2: The results of LO and NLO fits to HI & ZEUS data [31], with various lower cuts on Q², in the fit the number of flavors, f, is fixed to 4.

Q ² > 3 GeV ²	A ₀	A ₁	(Q ² /GeV ²) ^λ	f ² /total ²
LO	0.623±0.055	1.284±0.083	0.437±0.022	1.00
NLO _{min}	0.796±0.059	1.103±0.085	0.494±0.024	0.85
NLO _{max}	0.702±0.058	1.178±0.084	0.488±0.024	0.82
NLO	-0.352±0.041	1.335±0.080	0.700±0.044	1.05
NLO _{min}	0.192±0.046	1.029±0.086	0.517±0.060	0.74
NLO _{max}	-0.332±0.040	1.276±0.082	0.795±0.049	0.86
Q ² > 15 GeV ²				
LO	0.542±0.028	1.089±0.053	0.360±0.011	1.23
NLO _{min}	0.706±0.034	0.962±0.056	0.418±0.013	1.32
NLO _{max}	0.775±0.033	0.959±0.056	0.432±0.013	1.22
NLO	-0.109±0.021	1.248±0.053	0.538±0.023	1.82
NLO _{min}	0.116±0.024	0.967±0.064	0.499±0.030	1.04
NLO _{max}	0.139±0.022	1.087±0.061	0.478±0.028	1.27
Q ² > 2.5 GeV ²				
LO	0.526±0.023	1.049±0.045	0.352±0.009	1.87
NLO _{min}	0.760±0.028	0.919±0.048	0.422±0.010	1.48
NLO _{max}	0.798±0.028	0.908±0.047	0.428±0.010	1.90
NLO	-0.322±0.017	1.212±0.043	0.577±0.018	2.07
NLO _{min}	0.124±0.020	0.925±0.051	0.498±0.026	1.06
NLO _{max}	-0.153±0.018	1.016±0.051	0.488±0.021	1.51
Q ² > 0.5 GeV ²				
LO	0.566±0.014	1.053±0.036	0.298±0.005	5.24
NLO _{min}	0.865±0.012	0.864±0.039	0.318±0.006	3.13
NLO _{max}	0.874±0.012	0.875±0.021	0.368±0.006	2.96
NLO	-0.443±0.008	1.260±0.032	0.531±0.010	6.57
NLO _{min}	0.121±0.008	0.958±0.034	0.464±0.015	1.04
NLO _{max}	-0.071±0.007	0.712±0.023	0.520±0.011	2.39

	LO + NNLL	NLO _{min} + NNLL	NLO _{max} + NNLL	NLO + NNLL
σ _{tot} (Q ² /s ₀₎	24.31 ± 0.85	24.02 ± 0.36	24.17 ± 0.36	
σ _{had} (Q ² /s ₀)	15.40 ± 0.90	15.88 ± 0.37	15.89 ± 0.31	
Z _{had}	38.09	3.71	2.82	

Table 3: Fit results for σ_{tot}(Q²/s₀) and σ_{had}(Q²/s₀) at Q² = 80 GeV with 90% CL errors and minimum values of Z_{had} selected in the LO + NNLL, NLO_{min} + NNLL, and NLO_{max} + NNLL approximations.

Original	Traducción
Gloria: "Many roasts his own beans every Friday night!" Jay: "That kid is gonna be roasting his own beans for a long time, if you know what I mean!"	Gloria: "Many gusta su propio café los viernes por la noche" Jay: "Y va a seguir tostado su propio café mucho tiempo, si tu me entiendes"

131) Contexto
Claire llama a Phil para asegurarse que ha ido a buscar el regalo para la boda.

Original	Traducción
Chaire: "But you got the gift, right?" Phil: "That's what people tell me"	Chaire: "Pero tienes el regalo, no?" Phil: "Eso me dice la gente"

132) Contexto
Phil y Merle están relajándose en el balneario del club de golf de Jay. Ahí Merle conoce a uno de los amigos de Jay.

Original	Traducción
Jay: "He's visiting from Missouri" Amigo: "Uh, the 'Show-me' state" Jay: "Don't say that in here!"	Jay: "He venido de Missouri" Amigo: "Uh, el estado que le enseñan a leer" Jay: "No digas eso aquí dentro!"

133) Contexto
Andy va a Utah por aver a su novia, y está en el coche hablando con Haley.

Original	Traducción
Haley: "So, what are you gonna do on Utah?" Andy: "Well, my girl and I haven't seen eachother in a while, so we're gonna be like a couple of bunnys rabbits" Haley: "Waaaah!" Andy: "Just keeping all over town" Haley: "Oh"	Haley: "Y qué vas a hacer en Utah?" Andy: "Pues, no veo a mi novia desde hace tiempo, así que vamos a ser como un par de conejillos" Haley: "Voyaa" Andy: "Correteando por toda la ciudad" Haley: "Ah"

Branch: M.A. Political Science & Public Administration Time: 2.00pm to 5.00pm

Date & Day	Paper	Titles	Marks
11-10-2011	I	Political Thought	80
02-11-2011	II	Modern Political Theory	80
08-12-2011	III	Comparative Government & Politics	80
08-13-2011	IV	Public Policy	80
02-13-2011	V	Administrative Theories	80

Branch: M.A. History Time: 2.00pm to 5.00pm

Date & Day	Paper	Titles	Marks
11-10-2011	I	History of Modern Europe (1870-1912 A.D.)	80
02-11-2011	II	History of Modern India (1975-1956 A.D.)	80
08-12-2011	III	History of East Asia (China and Japan 1840-1910 A.D.)	80
08-13-2011	IV	Historical Method & Historiography	80
02-13-2011	V	Social and Cultural History of India up to 1206 A.D.	80

Branch: M.Sc., Mathematics Time: 2.00pm to 5.00pm

Date & Day	Paper	Titles	Marks
11-10-2011	I	Algebra	80
02-11-2011	II	Real Analysis	80
08-12-2011	III	Differential Equations	80
08-13-2011	IV	Computing Techniques (There is no Practical Examinations for this paper)	80
02-13-2011	V	Complex Analysis	80

Branch: M.A. Hindi Time: 2.00pm to 5.00pm

Date & Day	Paper	Titles	Marks
11-10-2011	I	Audhunk Hindi Kavya	80
02-11-2011	II	Hind Gadya Vidhayan	80
08-12-2011	III	Bhasha Vigyan aur Hind Bhasha ka Itihās	80
08-13-2011	IV	Hind Sahitya ka Itihās	80
02-13-2011	V	Telugu Bhasha aur Sahitya ka Itihās	80



Figure 5: Video-to-image translation results (Model1+Model2+Model3) and their reconstructed results (4)-(6) on the Celeb and PaSC databases

FaceNet Score	Image-to-Video	Video-to-Image
Model1	0.8011	0.8023
Model2	0.4991	0.4848
Model3	0.0428	0.0358

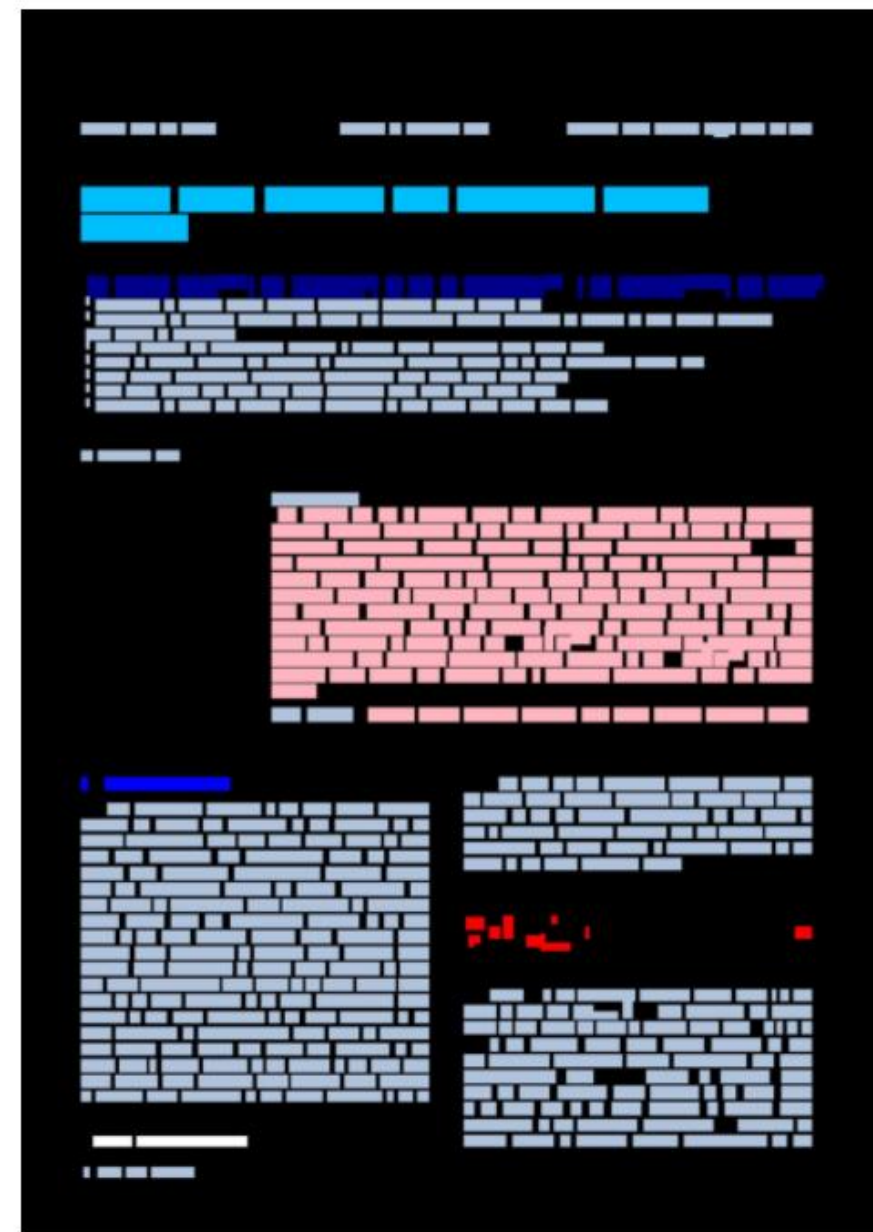
Table 1: FaceNet score (measured by L2 distance of the FaceNet [13] face embeddings) of image-to-video and video-to-image translation for the three proposed models on the Celeb and PaSC datasets.

In addition, we also study the identity score curve of the proposed Model3 that explicitly includes the identity preserving cost. As depicted in Fig. 8, on the two used datasets, the identity-based cost can always decrease stably as the training iteration number increases towards 50K. This demonstrates that the proposed model can be trained to transfer samples with keeping identity better.

5. Conclusion and Future Work
This paper presented a new problem of unpaired face translation between static images and dynamic videos which can be applied to video face prediction and enhancement. To handle the problem, we proposed a cyclic image-to-video translation model to bidirectionally translate between images and videos in a unified GAN model. To better preserve the identity during translation, we additionally introduce a face verification model that ensures to maintain the facial identity of the original input faces in images or videos. Evaluations on two standard face database

Final states	Benchmark	Backgrounds
$2\mu, k_{p1} \leq 3 [2\mu]$	15.62	31.84
$k_{p1} \leq 50 \text{ GeV}$	15.62	31.84
$k_{p1} - m_{p1} \geq 5 \text{ GeV}$	15.62	31.84
$k_{p1} - m_{p1} > 10 \text{ GeV}$	15.62	31.84
$k_{p1} - m_{p1} < 125 \text{ GeV}$	1.32	2.55
Significance	31.22	0.00
$k_{p2} - m_{p1} - m_{p2} \leq 10 \text{ GeV}$	15.62	15.92
Significance	2.47	2.51
$k_{p2} - m_{p1} - m_{p2} \leq 5 \text{ GeV}$	5.21	7.96
Significance	2.05	2.82

Table 4.12: The number of events for $m_{p1} \leq 2 [2\mu]$, $k_{p1} \leq 20 \text{$



MNRAS 000, 1–7 (2000) Preprint 23 November 2019 Compiled using MNRAS L^AT_EX style file v3.0

Cosmic String Detection with Tree-Based Machine Learning

A. Vafaei Sadr^{1,2,3}, M. Farhang¹, S. M. S. Movahed^{1,4}*, B. Bassett^{3,5,6,7}, M. Kunz²

¹ Department of Physics, Shahid Beheshti University, Velenjani, Tehran 19378, Iran
² Department of Physics, Theoretical and Cosmic for Astrophysics, University of Geneva, 12 Quai Ernest Ansermet, 1211 Geneva 4, Switzerland
³ Abruzzo Institute for Mathematical Sciences, 6 Molise Road, Montebelluna 36101, South Africa
⁴ School of Physics, Institute for Research in Fundamental Sciences (IPM), P. O. Box 19395-5531, Tehran, Iran
⁵ South African Astronomical Observatory, Observatory, Cape Town, 7935, South Africa
⁶ SCA, South Africa, The Park, Park Road, Pietermaritzburg, Cape Town 6001, South Africa
⁷ Department of Maths and Applied Maths, University of Cape Town, Cape Town, South Africa

23 November 2019

ABSTRACT
 We explore the use of random forest and gradient boosting, two powerful tree-based machine learning algorithms, for the detection of cosmic strings in maps of the cosmic microwave background (CMB), through their unique Gini-Kaiser-Selwyn effect on the temperature anisotropies. The information in the maps is compressed into feature vectors before being passed to the learning units. The feature vectors contain various statistical measures of processed CMB maps that boost the cosmic string detectability. Our proposed classifiers, after training, give results improved over or similar to the claimed detectability levels of the existing methods for string tension, $G\mu$. They can make 3 σ detection of strings with $G\mu \geq 2.1 \times 10^{-7}$ for noise-free, 3σ resolution CMB observations. The minimum detectable tension increases to $G\mu \geq 3.0 \times 10^{-8}$ for a more realistic, CMB SA-like (H) strategy, still a significant improvement over the previous results.

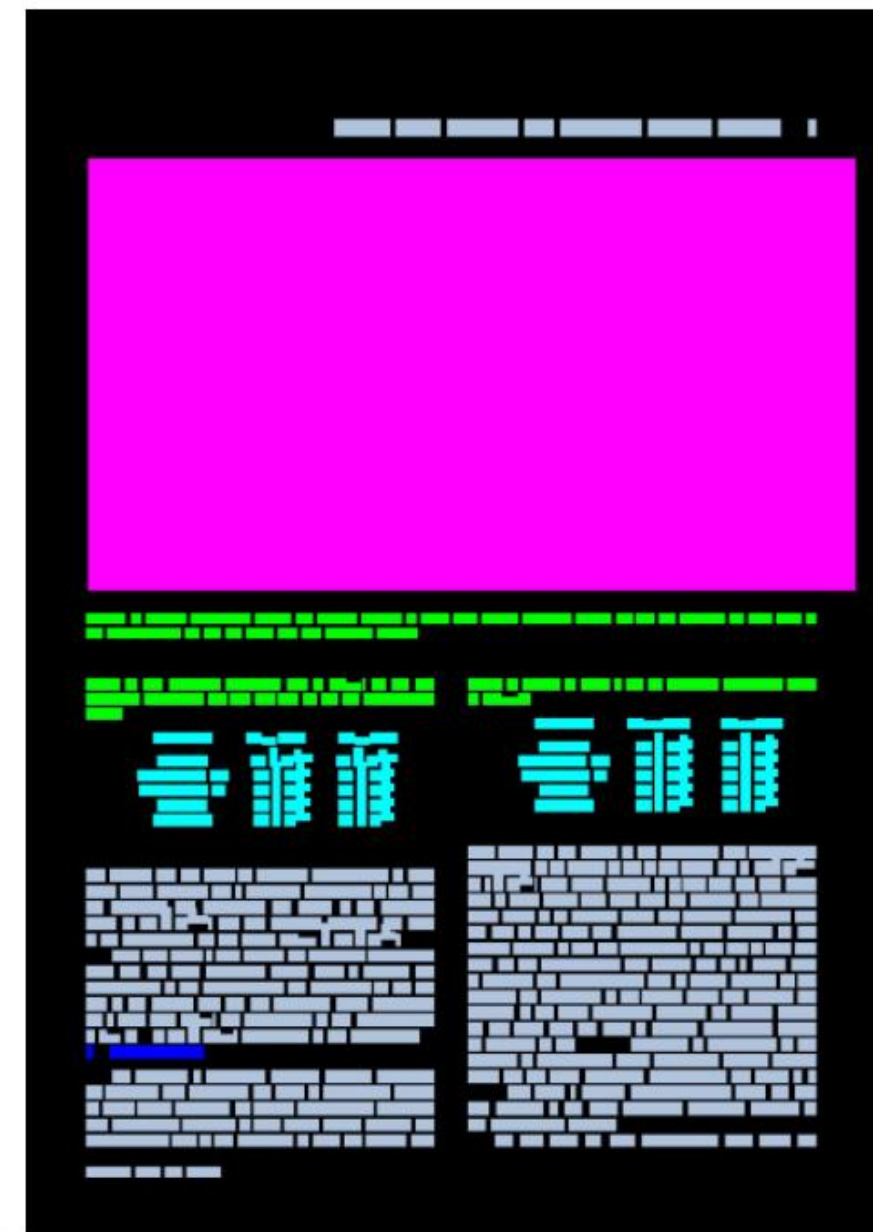
Key words: Cosmic string, Machine learning, Tree based models, CMB, CMB.

1 INTRODUCTION
 The inflationary paradigm is the most widely accepted scenario for seeding the structure in the Universe, as far as our observational tests with flying colors. There is, however, both theoretical and observational reasons for contributions from alternative, well-motivated scenarios. Among these are perturbations sourced by cosmic topological defects formed at cosmological phase transitions. In particular, cosmic strings (CS) are theoretically expected to be produced in the early Universe (Kibble 1976; Zeldovich 1968; Vilenkin 1981; Vachaspati & Vilenkin 1984; Vilenkin 1985; Shellard 1987; Hochstadt & Kibble 1995; Vilenkin & Shellard 2000; Subramanian 2007; Davis et al. 2008; Degen 2009; Davis et al. 2010; Copeland et al. 1984; Subramanian 1997; Sarangi & Tye 2002; Copeland et al. 2004; Pogosyan et al. 2004; Majumdar & Christian-Davies 2002; Dvali & Vilenkin 2004; Kibble 2004; Hory Tye 2008). The detection of CS would open a unique window to the physics of the early Universe (Kibble 1976; Zeldovich 1968; Vilenkin 1981; Vilenkin & Shellard 2000; Finocchia & Tye 2005). Therefore a lot of effort has been put into developing powerful statistical tools for cosmic string network detection and putting tight upper bounds on the CS tension, parametrized by $G\mu$, where G and μ represent Newton's constant and the string's tension, respectively. The string tension is intimately related to the energy of the phase transition epoch,

$$\frac{G\mu}{c^2} = \nu \left(\frac{m^2}{M_{\text{Planck}}^2} \right) \quad (1)$$

where ν is the symmetry breaking energy scale, c is the speed of light and $M_{\text{Planck}} = \sqrt{\hbar c/8\pi G}$ represents the Planck mass. In this paper we work in natural units with $\hbar = c = 1$. A CS network would leave unique imprints on cosmic microwave background (CMB) anisotropies, The Gini-Kaiser-Selwyn (KS) effect (Kaiser & Selwyn 1984; Gini 1931 1985; Selwyn 1988; Baecher et al. 1988; Allen et al. 1992; Fox et al. 1997; Singard & Baecher 2012) corresponds to the integrated Sachs-Wolfe effect caused by moving strings. It produces line-like discontinuities on the

* E-mail: s.movahed@ipm.ir
 © 0000 The Authors



Cosmic String Detection with Tree-Based Machine Learning 5

Figure 8. Feature importance reports the average number of times each feature appeared among the top ten features, for each type of the processor, for the RF (top) and GB (bottom) learner.

Figure 9. Feature importance reports the average number of times each feature appeared among the top ten features, for each type of the processor, for the RF (top) and GB (bottom) learner.

Table 1. The minimum detectable $G\mu$ or $G\mu_{95\%}$ for the two tree-based algorithms, GB and RF, and for the 3 σ experimental range.

experiment	$G\mu_{95\%}(\text{GB})$	$G\mu_{95\%}(\text{RF})$
noise-free	4.1×10^{-8}	2.1×10^{-8}
CMB-SA-like (H)	4.2×10^{-7}	3.0×10^{-7}
CMB-SA-like (B)	1.3×10^{-7}	1.0×10^{-7}
ACT-like	4.0×10^{-7}	1.0×10^{-7}
Planck-like	1.0×10^{-7}	1.0×10^{-7}

Table 2. Similar to Table 1, but for minimum detectable $G\mu$ s, of $G\mu_{95\%}$.

experiment	$G\mu_{95\%}(\text{GB})$	$G\mu_{95\%}(\text{RF})$
noise-free	3.8×10^{-9}	3.8×10^{-9}
CMB-SA-like (H)	1.2×10^{-7}	1.1×10^{-7}
CMB-SA-like (B)	4.2×10^{-7}	2.3×10^{-7}
ACT-like	5.0×10^{-7}	5.0×10^{-7}
Planck-like	1.0×10^{-7}	1.0×10^{-7}

old classes, can not make an unbiased measurement of each small string tension. For a non-detectable observation of the sky, the algorithm can distinguish the tension of CS (see also Davis et al. 2011), and can correctly estimate the level of CS contribution for $G\mu$ above $G\mu_{95\%} = 3.6 \times 10^{-7}$. Note that Table 1 only reports the minimum measurable $G\mu$ s and not their associated errors. That is because the uncertainties in our measurements are dominated by the size of $G\mu$ classes, and not the statistical error. Therefore, for a class with $G\mu_{95\%}$, the uncertainty in the measurement is $\delta G\mu = \Delta G\mu \times G\mu_{95\%}$, irrespective of the experiment.

7 DISCUSSION
 We proposed a tree-based machine learning algorithm for detecting and measuring the trace of CS-induced signals on CMB maps, simulated for noise observations, as realistic. Our simulations consisted of 1000 maps, passed through the pre-processing unit of the algorithm to form the feature vec-

tors, which are the inputs to the classifiers. The simulation consisted of 15 classes of $G\mu$ in the range $G\mu = 2.5 \times 10^{-7}$ to 4.0×10^{-7} , with equal spacing in $\ln G\mu$, and our null class. Out of these maps, 10% were used for training the classifiers (Gini takes as the random forest and gradient boosting) and the rest as test sets. We performed feature analysis on the feature vectors to find the significance of the role of each feature for the classification. The results can be a major help in choosing the important set of feature analysis by decreasing the dimension of the feature space and limiting the analysis to the most significant features. As general results we can state that the scale of cosine components should be included to the effective resolution of experiments in the presence of experimental noise, large-scale curl-like components are the most important components. For $G\mu$ it is difficult to make a definite recommendation, while the second moment is the most important statistical measure in the classification process.

We find that, for each experimental case, these $G\mu$

The colors of semantic structure labels are:

- Abstract
- Author
- Caption
- Equation
- Figure
- Footer
- List
- Paragraph
- Reference
- Section
- Table
- Title

ReadingBank

Children and young people with a clear mental health diagnosis
 OSCA will offer care to children and young people who have already been specialist CAMHS criteria, i.e. have been diagnosed with a serious mental health disorder where:

- The child/young person is at risk of placement breakdown and failed to engage with or disengaged from specialist CAMHS services
- Where the intensity of an intervention required to support a child in placement is greater than the resources available within specialist CAMHS (more than one visit per week required), and there is a history of the child and young person failing to engage with these services on a regular basis

Where a mental health diagnosis is less clear
 OSCA will care coordinate complex cases that meets at least one of the following criteria:

- The child/young person is looked after, adopted or under a child protection plan
- The young person is significantly involved in the criminal justice system / or has major substance misuse issues
- The young person has an Education, Health and Care Plan (EHCP), and is educated within specialist educational provision

And where:

- Significant emotional, behavioural, or mental health concerns that have been identified through the Youth Offending Service (YOS) team assessment, Education, Health and Care Plan (EHCP), Family Star assessment, or Department of Health Framework for Assessment

And where a minimum of two of the following criteria apply:

- The child or young person is at risk of placement breakdown (either home or a care placement)
- The child and young person's needs cannot be met by the range of professionals currently involved with the case
- A standard primary mental health intervention is CLEARLY not sufficient to meet the child's needs
- A range of other primary mental health interventions have already been tried and have proved unsuccessful or there is a history of failure to engage

Consultation and support to frontline professionals
 OSCA will provide named workers to support the following agencies:

- Looked after children's services
- Youth offending and substance misuse services
- Special schools for children with emotional difficulties

OSCA will provide support, advice and opportunities for frontline children services to ensure that children and young people are:

- Appropriately supported at the right level of care
- Ensure timely access into additional services where required

Child Interventions
 The OSCA team will provide Child support which involves intensive community-oriented treatment to children and young people in the acute / crisis phase of mental illness, which, in the absence of the Child Intervention, would be at risk of repeat admission.

OSCA team will

Page 3 of 8

(a)

Children and young people with a clear mental health diagnosis
 OSCA will offer care to children and young people who have already been specialist CAMHS criteria, i.e. have been diagnosed with a serious mental health disorder where:

- The child/young person is at risk of placement breakdown and failed to engage with or disengaged from specialist CAMHS services
- Where the intensity of an intervention required to support a child in placement is greater than the resources available within specialist CAMHS (more than one visit per week required), and there is a history of the child and young person failing to engage with these services on a regular basis

Where a mental health diagnosis is less clear
 OSCA will care coordinate complex cases that meets at least one of the following criteria:

- The child/young person is looked after, adopted or under a child protection plan
- The young person is significantly involved in the criminal justice system / or has major substance misuse issues
- The young person has an Education, Health and Care Plan (EHCP), and is educated within specialist educational provision

And where:

- Significant emotional, behavioural, or mental health concerns that have been identified through the Youth Offending Service (YOS) team assessment, Education, Health and Care Plan (EHCP), Family Star assessment, or Department of Health Framework for Assessment

And where a minimum of two of the following criteria apply:

- The child or young person is at risk of placement breakdown (either home or a care placement)
- The child and young person's needs cannot be met by the range of professionals currently involved with the case
- A standard primary mental health intervention is CLEARLY not sufficient to meet the child's needs
- A range of other primary mental health interventions have already been tried and have proved unsuccessful or there is a history of failure to engage

Consultation and support to frontline professionals
 OSCA will provide named workers to support the following agencies:

- Looked after children's services
- Youth offending and substance misuse services
- Special schools for children with emotional difficulties

OSCA will provide support, advice and opportunities for frontline children services to ensure that children and young people are:

- Appropriately supported at the right level of care
- Ensure timely access into additional services where required

Child Interventions
 The OSCA team will provide Child support which involves intensive community-oriented treatment to children and young people in the acute / crisis phase of mental illness, which, in the absence of the Child Intervention, would be at risk of repeat admission.

OSCA team will

Page 3 of 8

(b)

Table 5.8.32.A – Glasgow Point south neighbourhood plan: material change of use

Use	Categories of development and assessment	Assessment benchmarks
If in the neighbourhood plan area		
SCCA, if suitable development where not listed in this table	No change	Glasgow Point south neighbourhood plan code
If in the Alford use zone		
Centre activities (activity group)	Accepted development, subject to compliance with identified requirements	Not applicable
	If existing on existing premises, where:	
	(a) gross floor area is no greater than 1,500sqm for any individual tenancy where shop or shop component of a shopping centre;	
	(b) complying with all applicable outcomes in section 8 of the Centre or mixed use code	
	Accessible development – Code assessment:	
	If existing on existing premises, where:	Centre or mixed use code – purpose, overall outcomes and section 8 outcomes only
	(a) gross floor area is no greater than 1,500sqm for any individual tenancy where shop or shop component of a shopping centre;	
	(b) not complying with all applicable outcomes in section 8 of the Centre or mixed use code	
	If existing in new premises or an existing premises with an increase in gross floor area, where gross floor area is no greater than 1,500sqm for any individual tenancy where shop or shop component of a shopping centre	Glasgow Point south neighbourhood plan code Centre or mixed use code Prescribed secondary code

Page 3 of 8

(c)

Table 5.8.32.A – Glasgow Point south neighbourhood plan: material change of use

Use	Categories of development and assessment	Assessment benchmarks
If in the neighbourhood plan area		
SCCA, if suitable development where not listed in this table	No change	Glasgow Point south neighbourhood plan code
If in the Alford use zone		
Centre activities (activity group)	Accepted development, subject to compliance with identified requirements	Not applicable
	If existing on existing premises, where:	
	(a) gross floor area is no greater than 1,500sqm for any individual tenancy where shop or shop component of a shopping centre;	
	(b) complying with all applicable outcomes in section 8 of the Centre or mixed use code	
	Accessible development – Code assessment:	
	If existing on existing premises, where:	Centre or mixed use code – purpose, overall outcomes and section 8 outcomes only
	(a) gross floor area is no greater than 1,500sqm for any individual tenancy where shop or shop component of a shopping centre;	
	(b) not complying with all applicable outcomes in section 8 of the Centre or mixed use code	
	If existing in new premises or an existing premises with an increase in gross floor area, where gross floor area is no greater than 1,500sqm for any individual tenancy where shop or shop component of a shopping centre	Glasgow Point south neighbourhood plan code Centre or mixed use code Prescribed secondary code

Page 3 of 8

(d)



Form in Chinese: 开放式基金财产业务申请表. Includes fields for applicant information, company details, and a barcode at the bottom.

Chinese

Form in English: RESPECTED BROWN & WILKINSON INTERNATIONAL TOBACCO. Includes technical specifications for a cigarette and a barcode at the bottom.

English

Form in Italian: MODULO ADESIONE A POLIZZA RESPONSABILITÀ CIVILE. Includes personal and insurance details for ANNI LUCA.

Italian

Form in Portuguese: IDENTIFICAÇÃO DO COORDENADOR DO PROJETO. Includes contact information for Maria Jesus Pinto.

Portuguese

Form in German: Bewerbungsformular - Lehrstelle 2017. Includes applicant details for Hans Schmit and contact information for ANDRITZ HYDRO GmbH.

German

Form in Spanish: FORMULARIO DE SOLICITUD DE BECA. Includes personal information for Carmen Garcia and contact details for CEPB.

Spanish

Form in French: FORMULAIRE D'INSCRIPTION Programme Vendanges en France 2020. Includes registration details for a wine harvest program.

French

Form in Japanese: 登録用紙. Includes registration details for a research project in traditional Japanese painting.

Japanese



Applications



Applications

- Information Extraction
 - 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>
 - Multi-lingual Form Understanding (**XFUND**)
 - <https://aka.ms/xfund>
- Classification
 - Document Image Classification (**RVL-CDIP**)
 - <https://www.cs.cmu.edu/~aharley/rvl-cdip/>
- VQA
 - Document Visual Question Answering (**DocVQA**)
 - <https://rrc.cvc.uab.es/?ch=17>
- Layout Analysis
 - Table Detection (**TableBank**)
 - <https://aka.ms/tablebank>
 - Page Object Detection (**DocBank**)
 - <https://aka.ms/docbank>
 - Reading Order Detection (**ReadingBank**)
 - <https://aka.ms/readingbank>

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

Table 2: Entity-level F1 scores of the four entity extraction tasks: FUNSD, CORD, SROIE and 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%	-

Table 5: Classification accuracy on the RVL-CDIP dataset

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	-

Table 6: Average Normalized Levenshtein Similarity (ANLS) score on the DocVQA dataset (until 2020-12-24), “QG” denotes the data augmentation with the question generation dataset.

Table Detection

Models	Word			Latex			Word+Latex		
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
ResNeXt-101 (Word)	0.9496	0.8388	0.8908	0.9902	0.5948	0.7432	0.9594	0.7607	0.8486
ResNeXt-152 (Word)	0.9530	0.8829	0.9166	0.9808	0.6890	0.8094	0.9603	0.8209	0.8851
ResNeXt-101 (Latex)	0.8288	0.9395	0.8807	0.9854	0.9760	0.9807	0.8744	0.9512	0.9112
ResNeXt-152 (Latex)	0.8259	0.9562	0.8863	0.9867	0.9754	0.9810	0.8720	0.9624	0.9149
ResNeXt-101 (Word+Latex)	0.9557	0.8403	0.8943	0.9886	0.9694	0.9789	0.9670	0.8817	0.9224
ResNeXt-152 (Word+Latex)	0.9540	0.8639	0.9067	0.9885	0.9732	0.9808	0.9657	0.8989	0.9311

Table 2: Evaluation results on Word and Latex datasets with ResNeXt- $\{101,152\}$ as the backbone networks

Models	Precision	Recall	F1
ICDAR 2013 (train)	0.9748	0.7997	0.8786
UNLV	0.9185	0.9639	0.9406
Marmot	0.7692	0.9844	0.8636
DeepFigures	0.8527	0.9348	0.8918
TableBank (ResNeXt-152, Word)	0.9725	0.8528	0.9087
TableBank (ResNeXt-152, Latex)	0.9658	0.9594	0.9625
TableBank (ResNeXt-152, Word + Latex)	0.9635	0.9039	0.9328
Tesseract	0.9439	0.7144	0.8133
Camelot	0.9785	0.6856	0.8063

Table 3: Evaluation results on ICDAR 2013 dataset

Page Object Detection

Models	Abstract	Author	Caption	Equation	Figure	Footer	List	Paragraph	Reference	Section	Table	Title	Macro average
BERT _{BASE}	0.9294	0.8484	0.8629	0.8152	1.0000	0.7805	0.7133	0.9619	0.9310	0.9081	0.8296	0.9442	0.8770
RoBERTa _{BASE}	0.9288	0.8618	0.8944	0.8248	1.0000	0.8014	0.7353	0.9646	0.9341	0.9337	0.8389	0.9511	0.8891
LayoutLM _{BASE}	0.9816	0.8595	0.9597	0.8947	1.0000	0.8957	0.8948	0.9788	0.9338	0.9598	0.8633	0.9579	0.9316
BERT _{LARGE}	0.9286	0.8577	0.8650	0.8177	1.0000	0.7814	0.6960	0.9619	0.9284	0.9065	0.8320	0.9430	0.8765
RoBERTa _{LARGE}	0.9479	0.8724	0.9081	0.8370	1.0000	0.8392	0.7451	0.9665	0.9334	0.9407	0.8494	0.9461	0.8988
LayoutLM _{LARGE}	0.9784	0.8783	0.9556	0.8974	1.0000	0.9146	0.9004	0.9790	0.9332	0.9596	0.8679	0.9552	0.9350
X101	0.9717	0.8227	0.9435	0.8938	0.8812	0.9029	0.9051	0.9682	0.8798	0.9412	0.8353	0.9158	0.9051
X101+LayoutLM _{BASE}	0.9815	0.8907	0.9669	0.9430	0.9990	0.9292	0.9300	0.9843	0.9437	0.9664	0.8818	0.9575	0.9478
X101+LayoutLM _{LARGE}	0.9802	0.8964	0.9666	0.9440	0.9994	0.9352	0.9293	0.9844	0.9430	0.9670	0.8875	0.9531	0.9488

Table 4: The performance of BERT, RoBERTa, LayoutLM and Faster R-CNN on the DocBank test set.

Reading Order Detection

Method	Encoder	Avg. Page-level BLEU \uparrow	ARD \downarrow
Heuristic Method	-	0.6972	8.46
LayoutReader (text only)	BERT	0.8510	12.08
	UniLM	0.8765	10.65
LayoutReader (layout only)	LayoutLM (layout only)	0.9732	2.31
LayoutReader	LayoutLM	0.9819	1.75

Table 2: Evaluation results of the LayoutReader on the reading order detection task, where the source-side of training/testing data is in the left-to-right and top-to-bottom order

Method	Avg. Page-level BLEU \uparrow	ARD \downarrow
Heuristic Method	0.3391	13.61
Tesseract OCR	0.7532	1.42
LayoutReader	0.9360	0.27

Table 5: Adaption to text lines of Tesseract OCR

Method	Avg. Page-level BLEU \uparrow	ARD \downarrow
Heuristic Method	0.3752	10.17
The commercial OCR	0.8530	2.40
LayoutReader	0.9430	0.59

Table 6: Adaption to text lines of the commercial OCR



Multilingual
Document AI



Language-specific Fine-tuning

	Model	FUNSD	ZH	JA	ES	FR	IT	DE	PT	Avg.
SER	XLM-RoBERTa _{BASE}	0.667	0.8774	0.7761	0.6105	0.6743	0.6687	0.6814	0.6818	0.7047
	InfoXLM _{BASE}	0.6852	0.8868	0.7865	0.6230	0.7015	0.6751	0.7063	0.7008	0.7207
	LayoutXLM _{BASE}	0.794	0.8924	0.7921	0.7550	0.7902	0.8082	0.8222	0.7903	0.8056
	XLM-RoBERTa _{LARGE}	0.7074	0.8925	0.7817	0.6515	0.7170	0.7139	0.711	0.7241	0.7374
	InfoXLM _{LARGE}	0.7325	0.8955	0.7904	0.6740	0.7140	0.7152	0.7338	0.7212	0.7471
	LayoutXLM _{LARGE}	0.8225	0.9161	0.8033	0.7830	0.8098	0.8275	0.8361	0.8273	0.8282
RE	XLM-RoBERTa _{BASE}	0.2659	0.5105	0.5800	0.5295	0.4965	0.5305	0.5041	0.3982	0.4769
	InfoXLM _{BASE}	0.2920	0.5214	0.6000	0.5516	0.4913	0.5281	0.5262	0.4170	0.4910
	LayoutXLM _{BASE}	0.5483	0.7073	0.6963	0.6896	0.6353	0.6415	0.6551	0.5718	0.6432
	XLM-RoBERTa _{LARGE}	0.3473	0.6475	0.6798	0.6330	0.6080	0.6171	0.6189	0.5762	0.5910
	InfoXLM _{LARGE}	0.3679	0.6775	0.6604	0.6346	0.6096	0.6659	0.6057	0.5800	0.6002
	LayoutXLM _{LARGE}	0.6404	0.7888	0.7255	0.7666	0.7102	0.7691	0.6843	0.6796	0.7206

Table 2: Language-specific fine-tuning accuracy (F1) on the XFUND dataset (fine-tuning on X, testing on X), where “SER” denotes the semantic entity recognition and “RE” denotes the relation extraction.

SER: Semantic Entity Recognition (headers, keys, values)

RE: Relation extraction for key-value pairs

Zero-shot Transfer

	Model	FUNSD	ZH	JA	ES	FR	IT	DE	PT	Avg.	
SER	XLM-RoBERTa _{BASE}	0.667	0.4144	0.3023	0.3055	0.371	0.2767	0.3286	0.3936	0.3824	
	InfoXLM _{BASE}	0.6852	0.4408	0.3603	0.3102	0.4021	0.2880	0.3587	0.4502	0.4119	
	LayoutXLM _{BASE}	0.794	0.6019	0.4715	0.4565	0.5757	0.4846	0.5252	0.539	0.5561	
	XLM-RoBERTa _{LARGE}	0.7074	0.5205	0.3939	0.3627	0.4672	0.3398	0.418	0.4997	0.4637	
	InfoXLM _{LARGE}	0.7325	0.5536	0.4132	0.3689	0.4909	0.3598	0.4363	0.5126	0.4835	
	LayoutXLM _{LARGE}	0.8115	0.6007	0.425	0.425	0.519	0.476	0.535	0.517	0.5215	
RE	XLM-RoBERTa _{BASE}	1996	0.2276	0.2659	0.1601	0.2611	0.2440	0.2240	0.2374	0.2288	0.231
	InfoXLM _{BASE}	2049	0.2423	0.2920	0.2405	0.2851	0.2481	0.2454	0.2193	0.2027	0.231
	LayoutXLM _{BASE}	3685	0.4388	0.5483	0.4494	0.4408	0.4708	0.4416	0.4090	0.3820	0.431
	XLM-RoBERTa _{LARGE}	2333	0.2765	0.3473	0.2421	0.3037	0.2843	0.2897	0.2496	0.2617	0.271
	InfoXLM _{LARGE}	2554	0.3100	0.3679	0.3156	0.3364	0.3185	0.3189	0.2720	0.2953	0.301
	LayoutXLM _{LARGE}	4795	0.5487	0.6404	0.5531	0.5696	0.5780	0.5615	0.5184	0.4890	0.541

on X), where

Table 3: Zero-shot transfer accuracy (F1) on the XFUND dataset (fine-tuning on FUNSD, testing on X), where “SER” denotes the semantic entity recognition and “RE” denotes the relation extraction.

Multitask Learning

	Model	FUNSD	ZH	JA	ES	FR	IT	DE	PT	Avg.
SER	XLM-RoBERTa _{BASE}	0.6633	0.883	0.7786	0.6223	0.7035	0.6814	0.7146	0.6726	0.7149
	InfoXLM _{BASE}	0.6538	0.8741	0.7855	0.5979	0.7057	0.6826	0.7055	0.6796	0.7106
	LayoutXLM _{BASE}	0.7924	0.8973	0.7964	0.7798	0.8173	0.821	0.8322	0.8241	0.8201
	XLM-RoBERTa _{LARGE}	0.7151	0.8967	0.7828	0.6615	0.7407	0.7165	0.7431	0.7449	0.7502
	InfoXLM _{LARGE}	0.7246	0.8919	0.7998	0.6702	0.7376	0.7180	0.7523	0.7332	0.7534
	LayoutXLM _{LARGE}	0.8068	0.9155	0.8216	0.8055	0.8384	0.8372	0.853	0.8650	0.8429
RE	XLM-RoBERTa _{BASE}	0.3638	0.6797	0.6829	0.6828	0.6727	0.6937	0.6887	0.6082	0.6341
	InfoXLM _{BASE}	0.3699	0.6493	0.6473	0.6828	0.6831	0.6690	0.6384	0.5763	0.6145
	LayoutXLM _{BASE}	0.6671	0.8241	0.8142	0.8104	0.8221	0.8310	0.7854	0.7044	0.7823
	XLM-RoBERTa _{LARGE}	0.4246	0.7316	0.7350	0.7513	0.7532	0.7520	0.7111	0.6582	0.6896
	InfoXLM _{LARGE}	0.4543	0.7311	0.7510	0.7644	0.7549	0.7504	0.7356	0.6875	0.7037
	LayoutXLM _{LARGE}	0.7683	0.9000	0.8621	0.8592	0.8669	0.8675	0.8263	0.8160	0.8458

Table 4: Multitask fine-tuning accuracy (F1) on the XFUND dataset (fine-tuning on 8 languages all, testing on X), where “SER” denotes the semantic entity recognition and “RE” denotes the relation extraction.

Challenges in Document AI



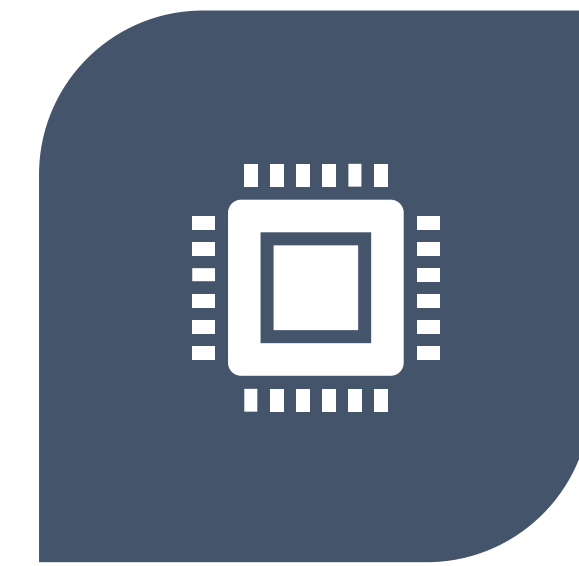
MODEL LIMITATIONS



DATA QUALITY IN
REAL-WORD



TASK CORRELATIONS



DATA/COMPUTATION
INSUFFICIENCY

Document AI @MSRA

- Multimodal Pre-trained Models

- **LayoutLM** (KDD'2020)
- **LayoutLMv2** (ACL'2021)
- **LayoutXLM** (Preprint)

- Benchmark Datasets

- **TableBank** (LREC'2020)
- **DocBank** (COLING'2020)
- **ReadingBank** (EMNLP'2021)
- **XFUND** (with LayoutXLM)

- Our paper **“Document AI: Benchmarks, Models and Applications”** will be publicly available soon

UniLM AI

Pre-trained (foundation) models across tasks (understanding, generation and translation), languages (100+ languages), and modalities (language, image, audio, vision + language, audio + language, etc.)

The family of UniLM AI:

UniLM (v1@NeurIPS'19 | v2@ICML'20 | v3@ACL'21): unified pre-training for language understanding and generation

InfoXLM (v1@NAACL'21 | v2@ACL'21): multilingual/cross-lingual pre-trained models for 100+ languages

DeltaLM (NEW): encoder-decoder pre-training for language generation and translation for 100+ languages

MiniLM (v1@NeurIPS'20 | v2@ACL'21): small and fast pre-trained models for language understanding and generation

AdaLM (v1@ACL'21): domain, language, and task adaptation of pre-trained models

LayoutLM (v1@KDD'20 | v2@ACL'21): multimodal (text + layout/format + image) pre-training for **Document AI** (e.g. scanned documents, PDF, etc.)

LayoutXLM (NEW): multimodal (text + layout/format + image) pre-training for multilingual document understanding

LayoutReader (EMNLP'21): Pre-training of text and layout for reading order detection

BEiT (NEW): BERT Pre-Training of Image Transformers

UniSpeech (v1@ICML'21): Speech Pre-Training for ASR and TTS

s2s-ft: sequence-to-sequence fine-tuning toolkit
<https://github.com/microsoft/unilm>

XLM-T (NEW): Multilingual NMT w/ pretrained cross-lingual encoders