

DUBLIN: Visual Document Understanding By using Language-Image Network



Kriti Aggarwal*, Aditi Khandelwal*, Kumar Tanmay*, Owais Mohammed Khan, Qiang Liu, Monojit Choudhury, Hardik Hansrajbhai Chauhan, Subhojit Som, Vishrav Chaudhary, Saurabh Tiwary



Introduction

TEAM	PLAYED	WON	LOST	NR	NET RR	POINTS
GT	14	10	4	0	+0.809	20
CSK	14	8	5	1	+0.652	17
LSG	14	8	5	1	+0.284	17
MI	14	8	6	0	-0.044	16
RR	14	7	7	0	+0.148	14
RCB	14	7	7	0	+0.135	14
KKR	14	6	8	0	-0.239	12
PKS	14	6	8	0	-0.304	12
DC	14	5	9	0	-0.808	10
SRH	14	4	10	0	-0.590	8

The IPL Points Table is one of the most challenging points tables in cricket as it can declare the team's fate for the qualifications. The tournament will begin with league matches, where each team will face the other twice. Then the top-4 teams will qualify for the knock-out games and then the top-2 will play the final. The IPL Points Table will come in handy at the end of the league stages when teams will look at their wins and losses. If a team wins on a regular basis, then there won't be any issue, but when a team loses their important matches in the group stages then their qualification will be totally based on the net run-rate (NRR). If the NRR is good then a team can qualify for the knockout stage but if the team has moderate or negative NRR, then the team could be in deep trouble. All 10-teams will get 2 points if they win a game. If a team loses, they won't get any points and it might affect their NRR. If the game is tied then it will be decided by a super-over which would decide the game. The knock-out games and the final would decide the fate of whether there would be a new champion or the defending champion retains the title.

Question 1: What is the Net RR of Royal Challengers Bangalore?
Answer: +0.135

Question 2: How is the game decided in case of tie?
Answer: It will be decided by a super-over.

Question 3: Which team scores the least points?
Answer: Sunrisers Hyderabad

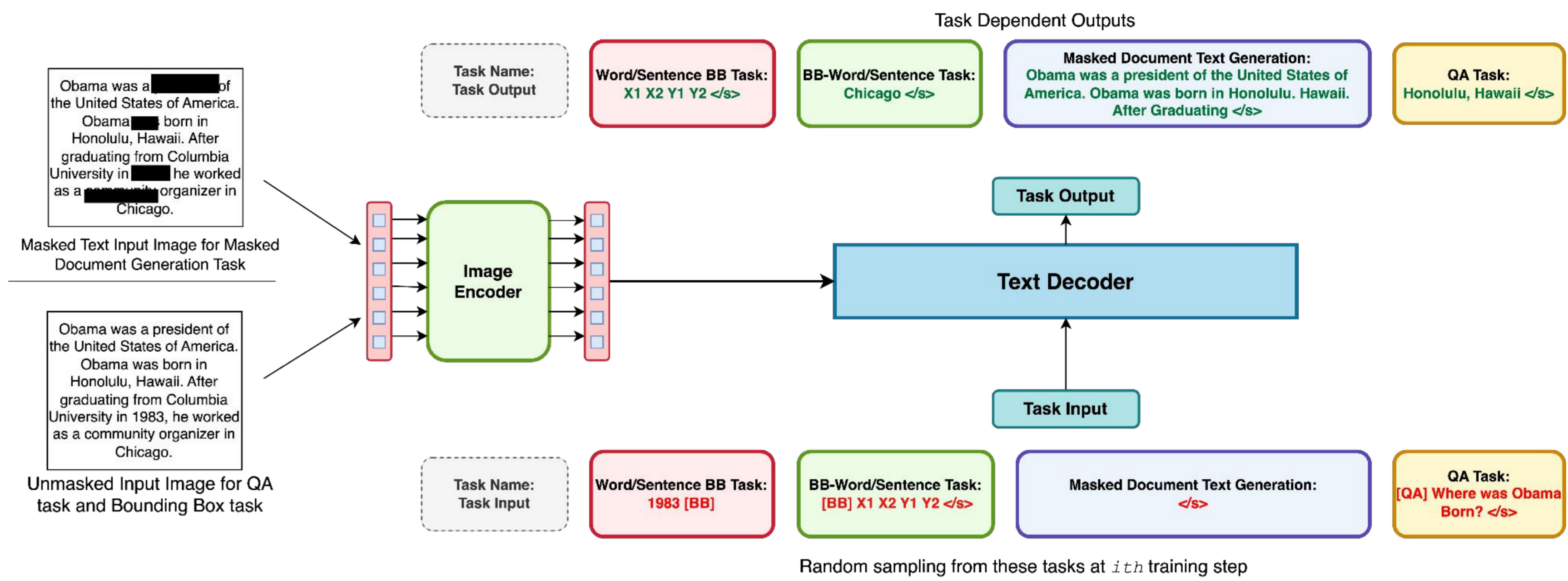
Document Understanding Challenges with Current Approaches:

- Limitations of OCR: Computationally expensive, error-prone, struggles with rich visual cues.
- Pixel-only models (e.g., Donut and Pix2struct) lack versatility for diverse tasks.

Features of DULIN

- OCR-free Visual Document Model
- Compact in size: it's only 976M parameter leveraging Bletchley-InfoXLM architecture
- Robust and Versatile: Handles diverse document images such as infographics, charts, forms, tables, natural images, webpages, UI, plain-text.
- Achieves SOTA on multiple tasks such as QA over illustrations, information extraction, table QA, UI understanding, image captioning etc.

Pre-training Framework



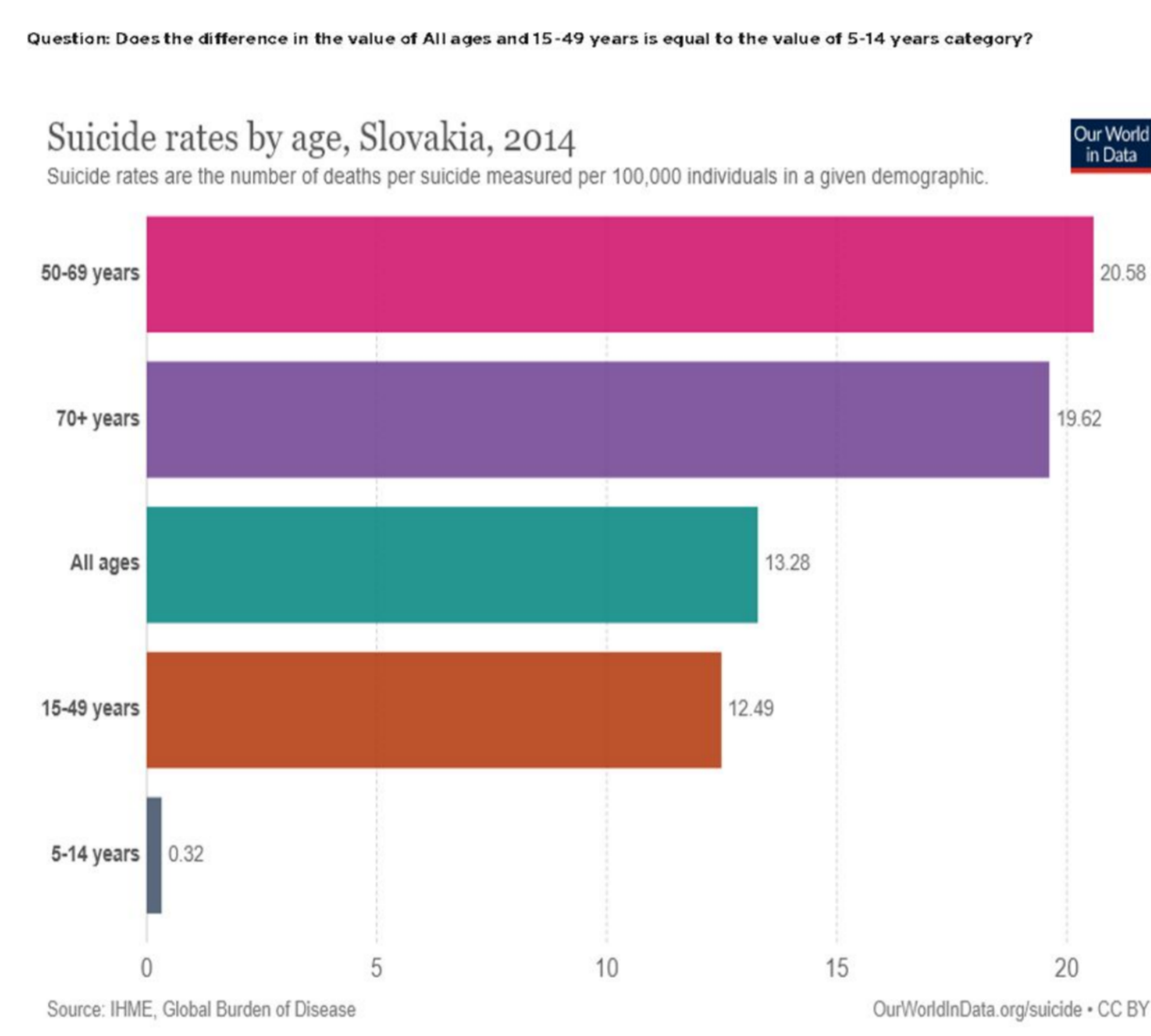
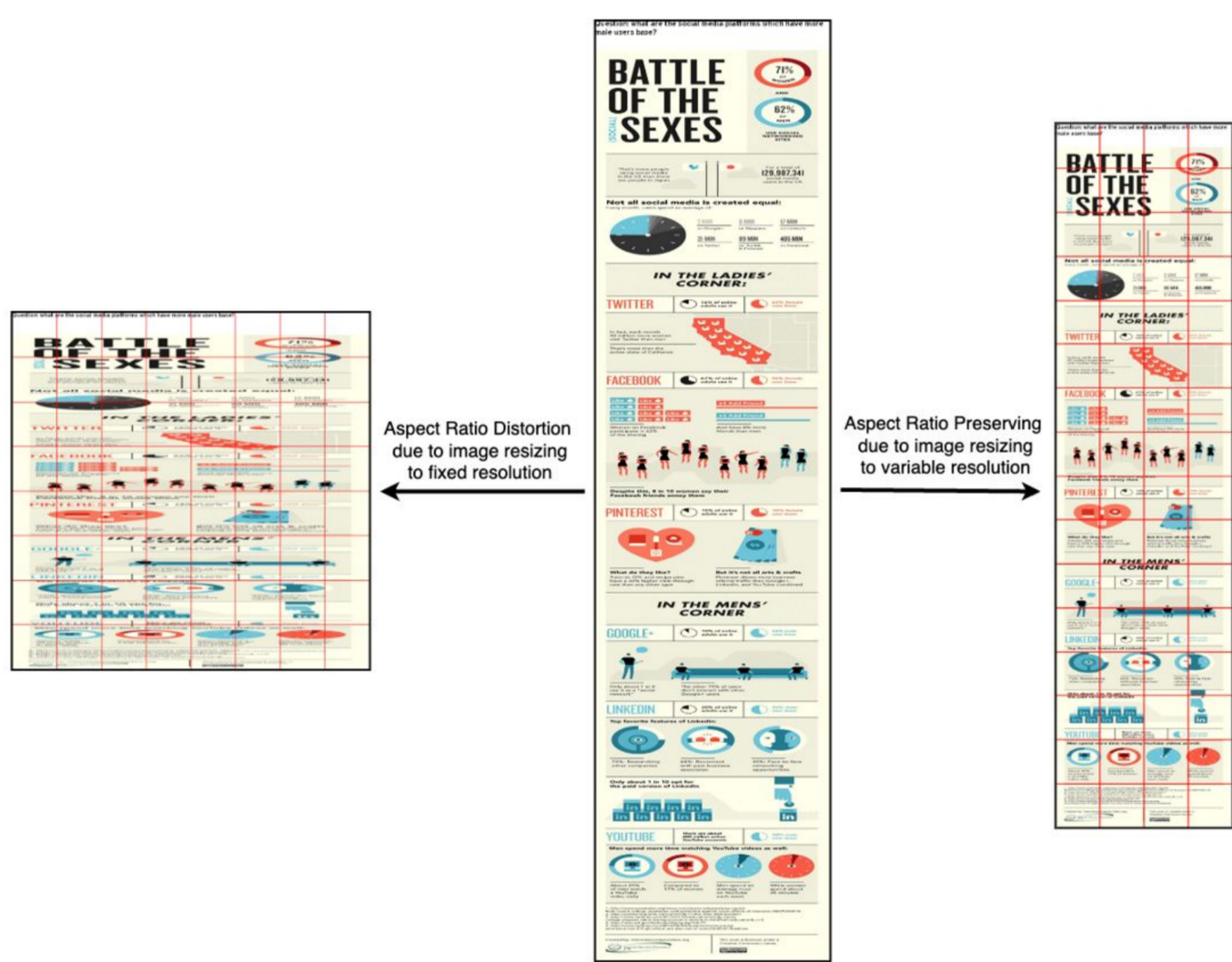
Pre-training Data	Pre-training Tasks
<ul style="list-style-type: none"> CCNews 200M Google NQ Dataset Rendered InfoXLM Dataset Bing QA Dataset Synthetic Table Structure QA Dataset 	<ul style="list-style-type: none"> Masked Document Language Modeling Task Bounding Box Task Rendered QA Task Masked Autoencoding (MAE) Task

Data Preprocessing Strategies before Finetuning

MAE Task

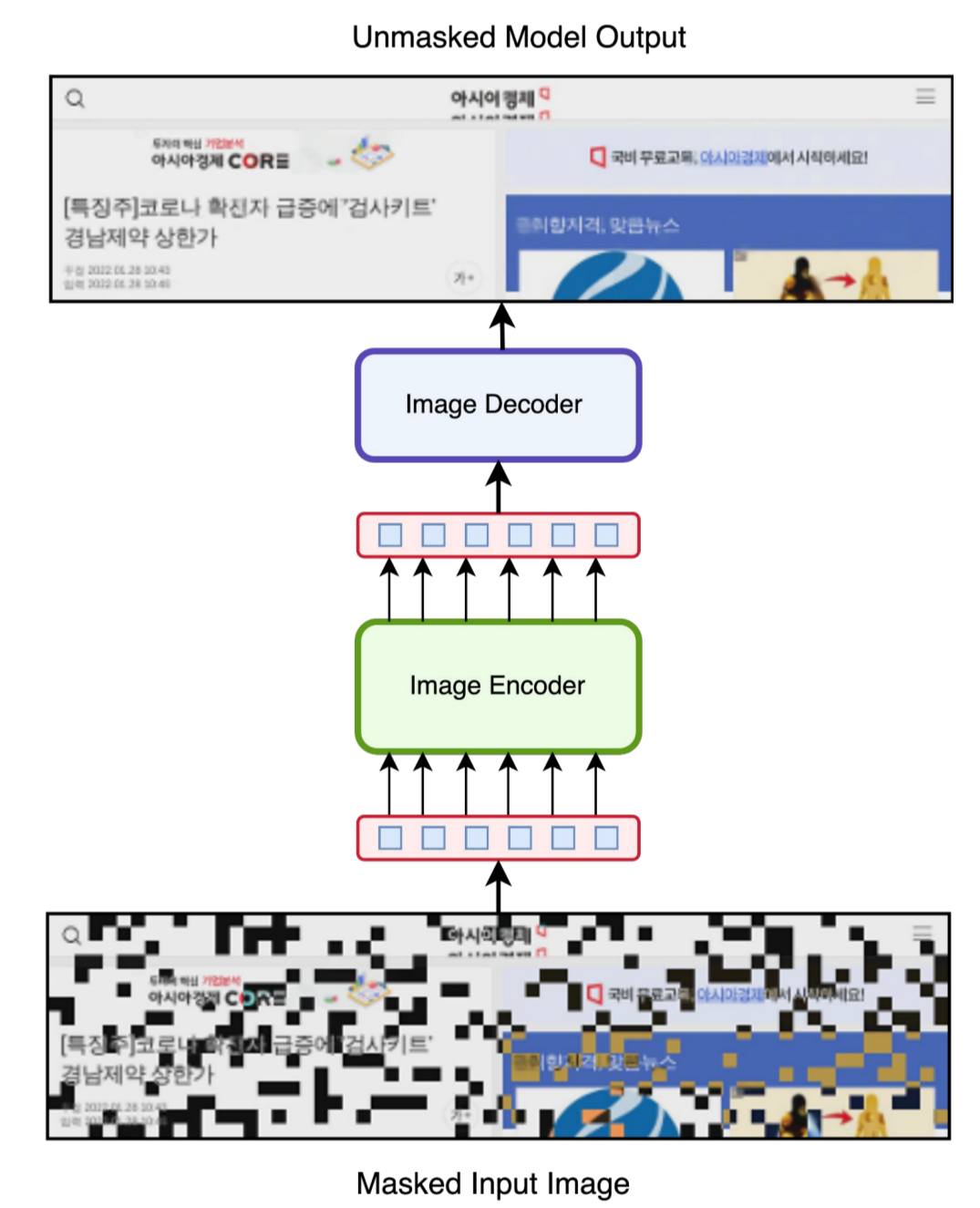
Efficient Variable Input Resolution Technique

Question on top of image

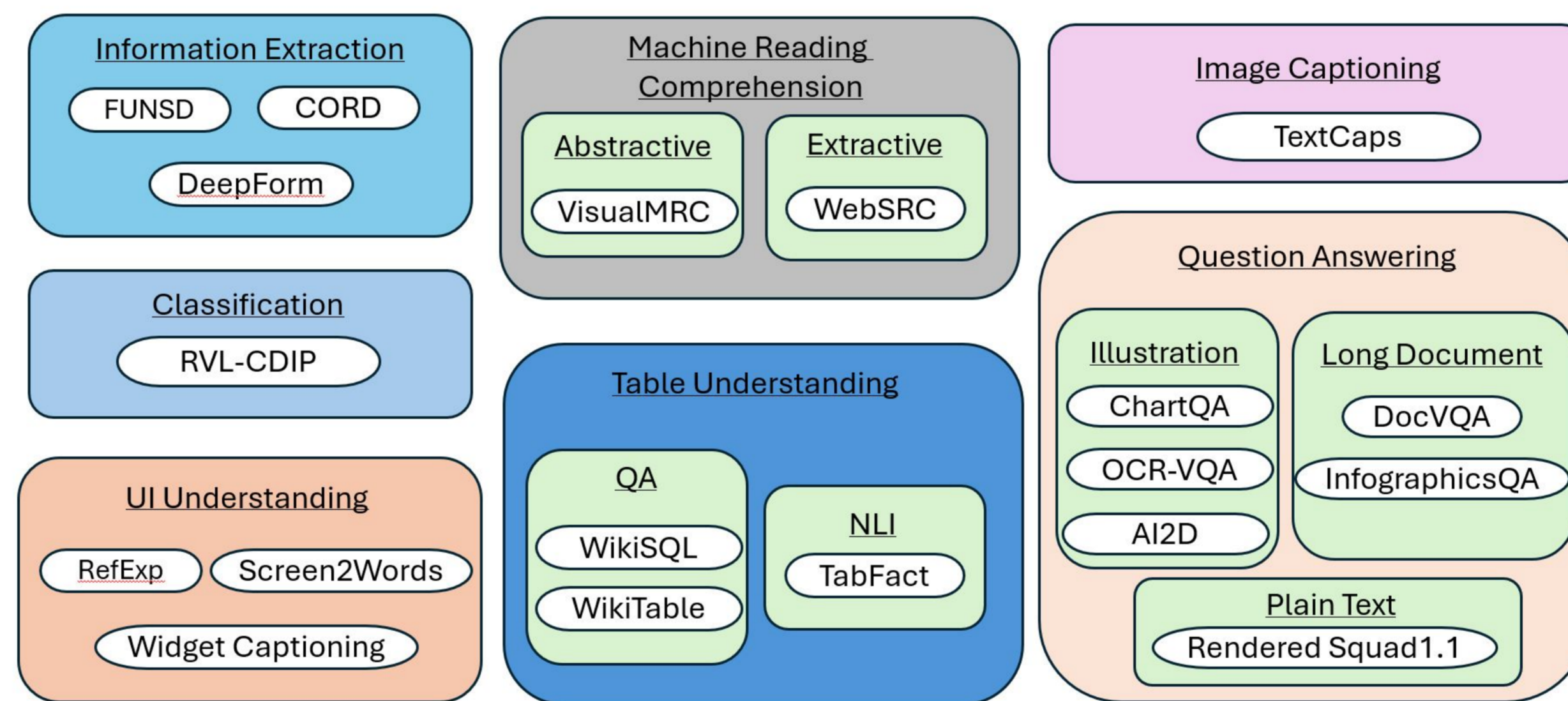


- Placing question on top of images with different font style and size to boost contextual understanding
- On an average, it leads to improvement by 3% over all datasets.

- Template based finetuning helps in executing diverse tasks without adding any external layers for specific tasks
- Leading to generalisation and achieving comparable performances to specialized pipeline models.



Template based fine tuning



Synthetic Table QA Dataset

Template	Example
What is the cell value in row [row_number] and column [column_number]?	What is the cell value in row 3 and column 2?
What is the cell value in column [column_number] and row [row_number]?	What is the cell value in column 7 and row 2?
What does the cell in the row [row_number] and column [column_number] contain?	What does the cell in row 4 and column 9 contain?
What does the cell in column [column_number] and row [row_number] contain?	What does the cell in column 1 and row 3 contain?
What is the cell value in column [column_name] and row [row_number]?	What is the cell value in column "Price" and row 4?
What is the value of cell where column is [column_name] and row number is [row_number]?	What is the value of cell where column is "Address" and row number is 9?
What is the value in the cell in [column ordinal] column where the row contains [row entry]?	What is the value in the cell in second column where the row contains "Mangoes"?
What is the value for [column 1 entries]?	What is the value for "City"?
How many rows are there in this table?	-
How many columns are there in this table?	-
What is the caption of the table?	-

Results and Analysis

Performance on QA over illustrations, UI understanding, image captioning, QA tasks

Model	QA over Illustrations			UI understanding			Captioning	Document QA	
	ChQA	AI2D	O-VQA	RefExp	Widget Cap	Scrnl2Wds	TCaps	DVQA	IVQA
Metrics	RA	ANLS	F1	EM	CIDEr	CIDEr	CIDEr	ANLS	ANLS
Donut	41.8	30.8	66.0	-	127.4	56.4	74.4	67.5	11.6
Pix2Struct _{large}	58.6	42.1	71.3	94.2	136.7	109.4	95.5	76.6	40.0
Dublin _{fixed_res}	35.6	51.1	73.1	99.1	132.2	101.8	92.8	78.2	36.8
Dublin _{variable_res}	35.2	52.3	74.0	99.1	132.2	101.8	92.8	80.7	43.0
(SOTA with spl. pipelines)	(VTP)	(DQAN)	(LATr)	(UIB)	(VUT)	(VUT)	(PaLI)	(UDOP)	(UDOP)
	45.5	38.5	67.5	90.8	97.0	64.3	160.4	84.7	47.4

Performance on Information Extraction, Classification and Reading Comprehension

Model	Information Extraction			Classification	Reading Comprehension	
	FUNSD	CORD	DeepForm	RVL-CDIP	WebSRC	VisualMRC
Metrics	F1	F1	F1	Accuracy	EM/F1	CIDEr
Donut	-	91.6	-	95.3	-	-
Dublin _{fixed_res}	77.8	97.1	62.2	94.9	77.7/84.2	347.3
Dublin _{variable_res}	77.8	97.1	65.7	94.9	77.7/84.2	347.3
SOTA with Spl. Pipelines	(LyLMv3)	(UDOP)	(UDOP)	(UDOP)	(TIE)	(LyT5-large)
	92.08	97.6	85.5	96.00	81.6/86.2	344.1
BERT _{large} /T5 (Text Baseline)	65.63	90.25	74.4	89.92	-	-

Performance on Table QA/NLI

Model	Table QA/NLI		
	WTQ	TabFact	WikiSQL
Metrics	EM	Accuracy	EM
Dublin _{fixed_res}	25.7	73.54	75.3
Dublin _{variable_res}	29.7	72.9	75.3
(SOTA w/ Spl. pipelines)	(UDOP)	(UDOP)	(TAPEX)
	47.2	78.9	89.2
(BART) Text Baseline	38.0	76.0	85.8

Date ⁽¹⁾	Rank	Tournament name	Venue	City	Winner	Runner-up	Score ⁽²⁾	Reference
09-09	WR	Asian Classic	Riverside Montien Hotel Civic Centre	Bangkok	Ronnie O'Sullivan	Brian Morgan	9-8	[1219]
09-24	WR	Scottish Masters	Edinburgh	Edinburgh	Peter Ebdon	Alan McManis	9-6	[181]
10-05	WR	Benson & Hedges Championship	JP Snooker Centre	Edinburgh	Brian Morgan	Drew Henry	9-8	[152]
10-08	WR	Malta Grand Prix	Jerma Palace Hotel	Marsaxlokk	Nigel Bond	Tony Drago	7-3	[161]
10-16	WR	Grand Prix	Bournemouth International Centre	Bournemouth	Mark Williams	Euan Henderson	9-5	[171]
10-29	WR	World Cup	Ameri Watergate Hotel	Bangkok	Scotland	John Higgins	10-7	[161]
11-15	WR	UK Championship	Crucible Theatre	Sheffield	Stephen Hendry	John Higgins	10-9	[161]
12-09	WR	German Open	NAHF	Düsseldorf	Ronnie O'Sullivan	Alan Robb	9-7	[160]
01-02	WR	Charity Challenge	International Convention Centre	Birmingham	Stephen Hendry	Ronnie O'Sullivan	9-8	[111]
01-24	WR	Welsh Open	Newport Leisure Centre	Newport	Stephen Hendry	Mark King	9-2	[121]
02-02	WR	Masters	Wandey Conference Centre	London	Steve Davis	Ronnie O'Sullivan	10-8	[13194]
02-13	WR	International Open	A.E.C.C.	Aberdeen	Stephen Hendry	Tony Drago	9-1	[13194]
02-23	WR	European Open	Mediteranean Conference Centre	Vallarta	John Higgins	John Parrott	9-5	[11194]
03-10	WR	Thai Open	Century Park Hotel	Bangkok	Peter Ebdon	Jiger Bond	9-7	[13194]
03-18	WR	Irish Masters	Gallop	Kill	Stephen Hendry	Brian Morgan	9-8	[12194]
03-27	WR	British Open	Plymouth Pavilions	Plymouth	Mark Williams	Stephen Hendry	9-2	[194]
04-19	WR	World Snooker Championship	Crucible Theatre	Sheffield	Ken Doherty	Stephen Hendry	18-12	[194]
05-27	WR	Profesionl	Proline	Proline	Martin Clark	Andy Hicks	9-7	[194]
12-28	WR	European League	Diamond Centre	Ilkinston	Ronnie O'Sullivan	Stephen Hendry	10-8	[121]

Question: What is the name of the first venue on this list?
DUBLIN's Answer: Riverside Montien Hotel
Gold Answer: Riverside Montien Hotel

Modou Bamba Gaye
From Wikipedia, the free encyclopedia

Modou Bamba Gaye is a Gambian politician who was the National Assembly Member for Lower Saloum, representing the National Reconciliation Party (NRP), from a 2015 by-election to the 2017 parliamentary election.

Political career [edit]

Gaye was elected at a 2015 by-election for the seat of Lower Saloum, following the dismissal of incumbent NAM Pa Malick Ceesay from the ruling Alliance for Patriotic Reorientation and Construction (APRC). Gaye defeated APRC candidate Kebba Touray in the election, winning 2764 votes to Touray's 1616 votes.^[1] Speaking in the National Assembly in January 2017, during the constitutional crisis and Yahya Jammeh's refusal to step down, Gaye called for a peaceful transition of power and said, "The people who voted us in are the same people who voted for Jammeh before and are the same people who voted Adama Barrow."^[2]

Question: When was Gaye elected for the seat of Lower Saloum?
DUBLIN's Answer: Gaye was elected at a 2015 by-election.
Gold Answer: In 2015

- DUBLIN is the first pixel only baseline on FUNSD, Deepform, WebSRC, VisualMRC, WTQ, and TabFact, WikiSQL and Squad1.1.
- We also introduce a new evaluation setup on text-based datasets by rendering them as images.
- It can be used in multiple applications from search engines to presentations.
- Possible future direction: Integrating Generative models like T-NLG or Llama models.

Contact: vchaudhary@microsoft.com

Link to paper: <https://arxiv.org/pdf/2305.14218.pdf>