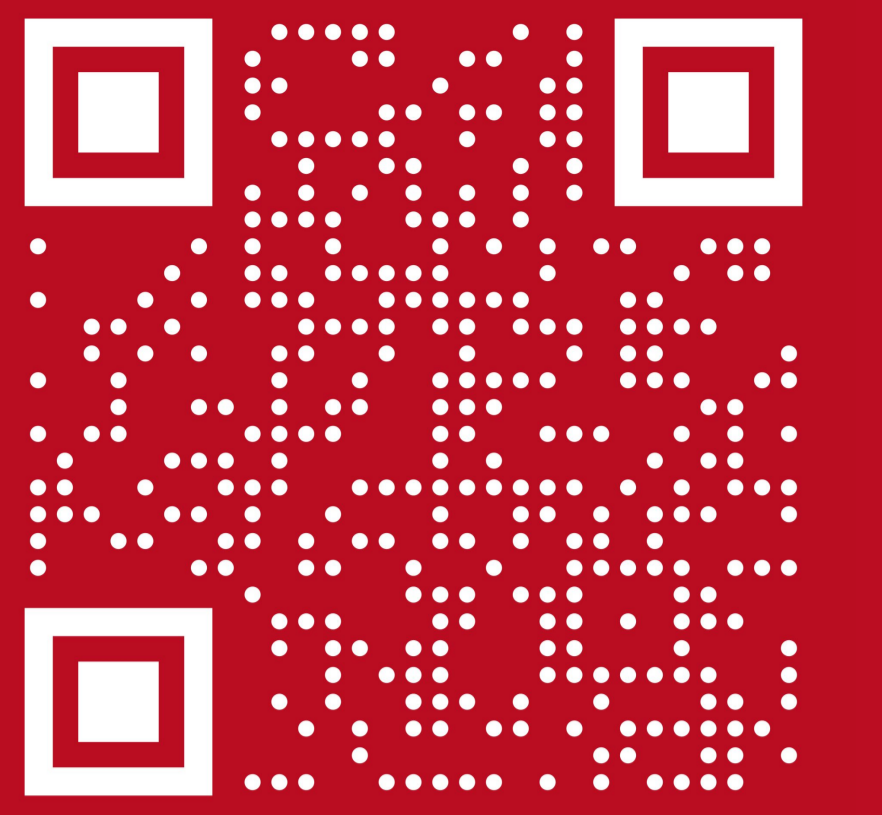


COFAR: Commonsense and Factual Reasoning in Image Search



COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

Prajwal Gatti¹, Abhirama Subramanyam Penamakuri¹, Revant Teotia², Anand Mishra¹,
Shubhashis Sengupta³, Roshni Ramnani³

¹IIT Jodhpur, India ²Columbia University ³Accenture Labs



GOAL

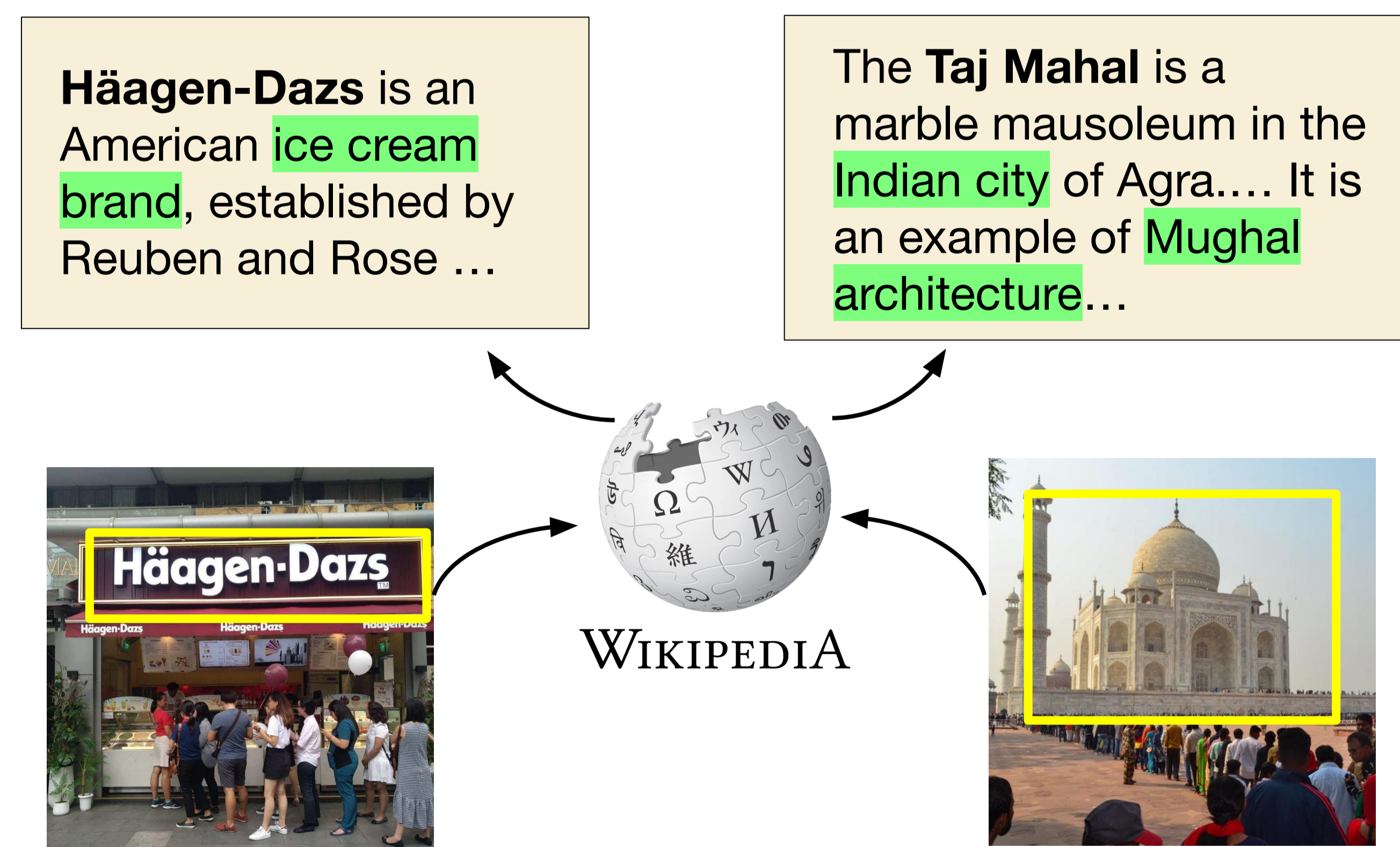
A queue of customers patiently waiting to buy ice cream



But...
Who is selling **ice cream**?
Who are **customers**?

Proposed Task:
Enable Commonsense and Factual Reasoning in Image Search

VISUAL NAMED ENTITY LINKING



Use named entities in images as a gateway to world knowledge: **Recognise** and **link to Wikipedia**

COFAR DATASET

A **new** dataset for studying our proposed task! Contains **40K+** queries, **25K+** images and **5K+** named entities



"The captain of the Argentina national football team celebrating after scoring a goal"

Entity: **Lionel Messi**
Type: **Celebrity**

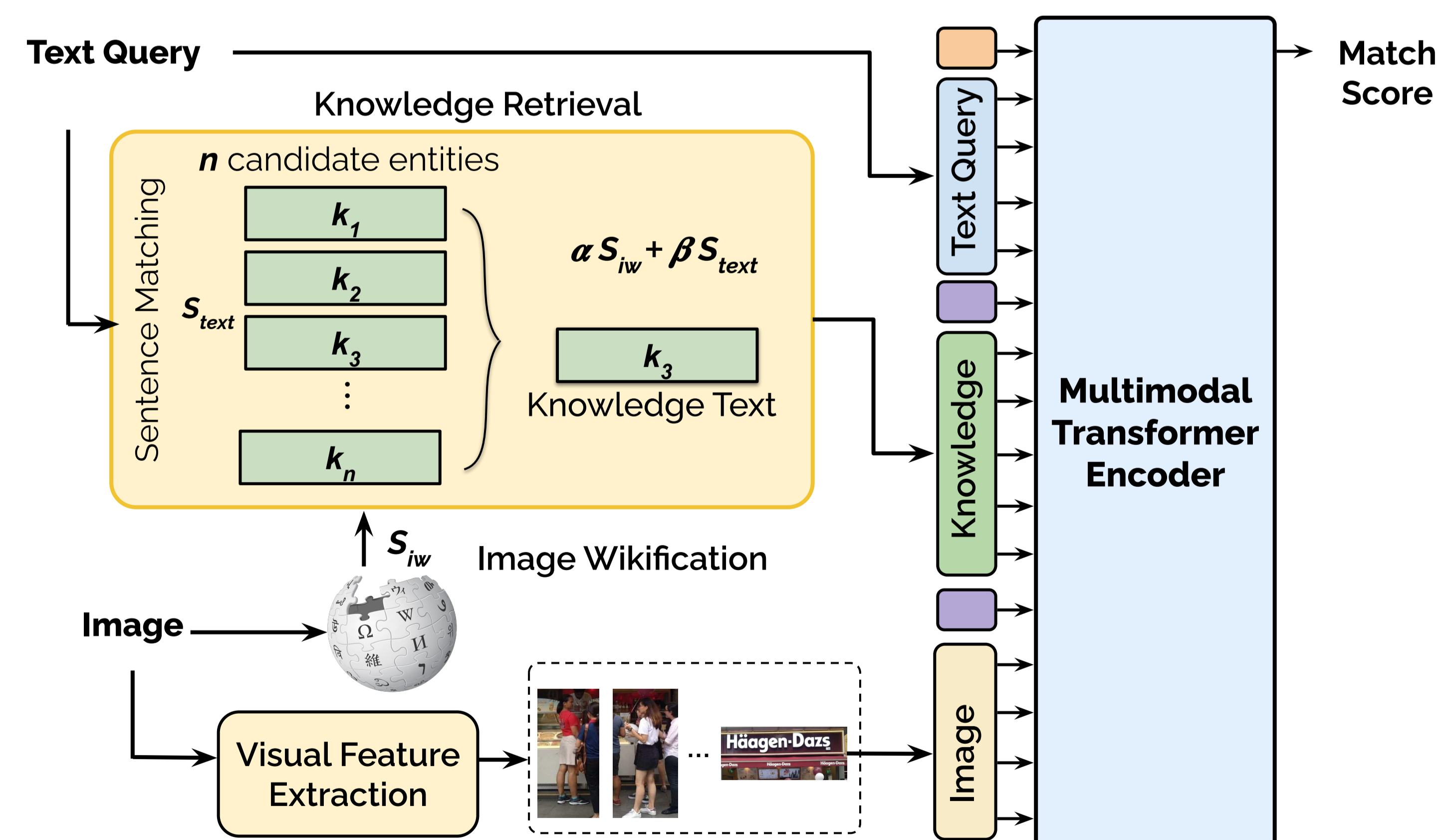
"A gray car waiting to refuel at a gas station"

Entity: **Petro Canada**
Type: **Business Brand**

"Two people getting married in front of a tower in Paris"

Entity: **Eiffel Tower**
Type: **Landmark**

KNOWLEDGE AWARE RETRIEVAL



We introduce **KRAMT**: A **K**nowledge **R**etrieval **A**ugmented **M**ultimodal **T**ransformer that retrieves relevant knowledge based on text and visual similarity

RESULTS

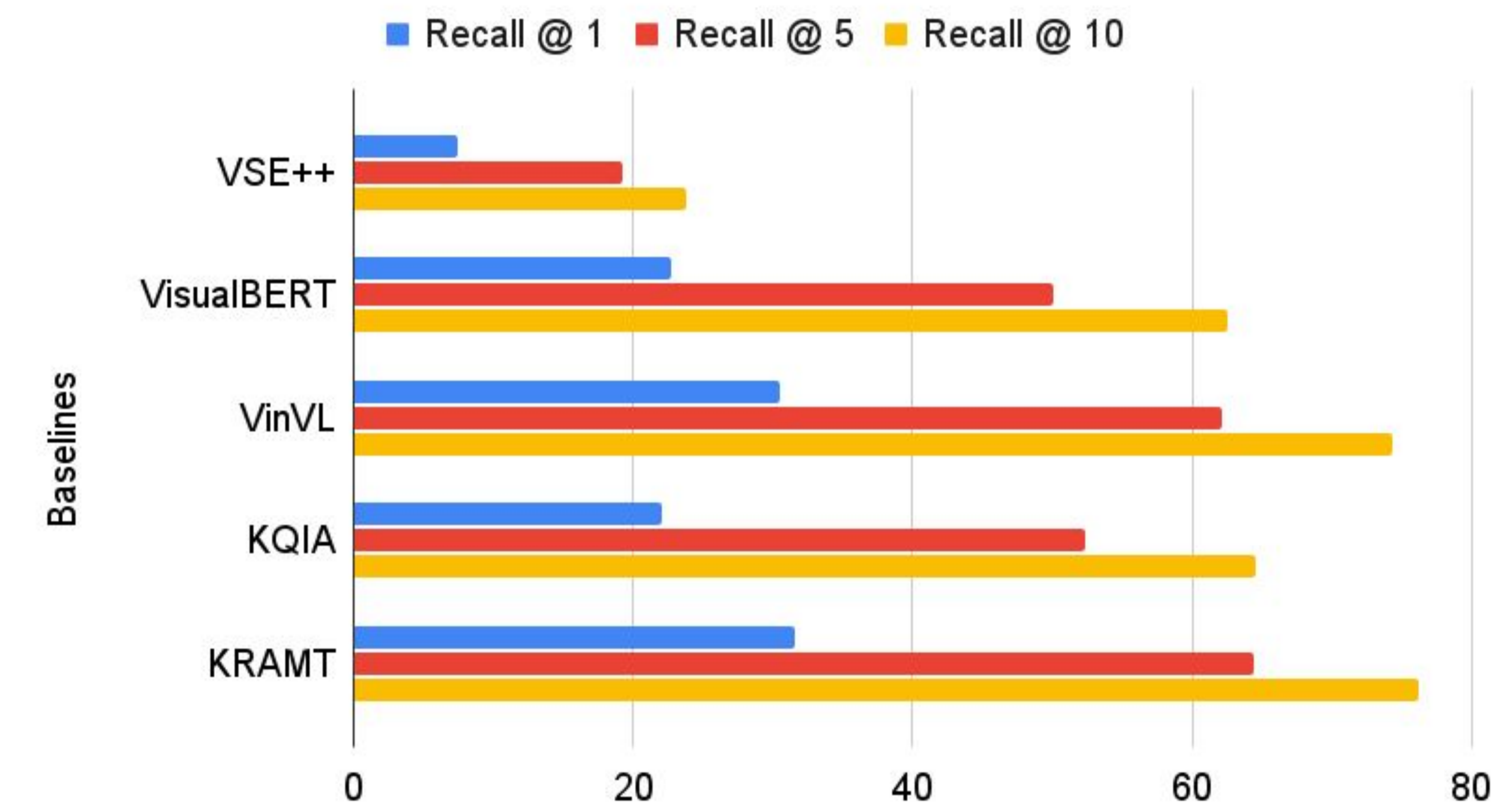
Query: "Lady handling a financial transaction inside a bank"



Query: "People protesting outside the world's most visited museum"



Results on COFAR-1K



SUMMARY

We proposed the novel task of image search that requires commonsense and factual reasoning

We provided a new human annotated dataset to study this task

We introduced a knowledge retrieval augmented method that uses named entities in images to reason with world knowledge

Acknowledgements: We thank Accenture Labs for supporting this work.