AACL IJCNLP 2022

COFAR: Commonsense and Factual Reasoning in Image Search

Prajwal Gatti, Abhirama S. Penamakuri,
Revant Teotia, Anand Mishra,
Shubhashis Sengupta, Roshni Ramnani

Columbia University
in the City of New York

भारतीय प्रौद्योगिकी संस्थान जोधपुर
Indian Institute of Technology Jodhpur
॥ त्वं ज्ञानमयो विज्ञानमयोऽसि ॥

accenture

Can **linking images to encyclopedic knowledge** enable commonsense and factual reasoning in image search?

Problem

A queue of people

॥ त्वं ज्ञानमयो विज्ञानमयोऽसि ॥

Problem

A queue of people

...

...

[Johnson et al., CVPR'15], [Faghri et al., BMVC'18], [Zhang et al., CVPR'19]

Problem

A queue of customers patiently waiting to buy ice cream

Who are customers?
Who is selling ice cream?

Problem

A queue of customers patiently waiting to buy ice cream

...

...

Who are customers?
Who is selling ice cream?

Problem

A queue of customers patiently waiting to buy ice cream

...

...

Who are customers?
Who is selling ice cream?

Problem

A queue of customers patiently waiting to buy ice cream

...

...

WIKIPEDIA

Who are customers?
Who is selling ice cream?

Problem

A queue of customers patiently waiting to buy ice cream

"Häagen-Dazs is an American ice cream brand …"

"Starbucks Corporation is a chain of coffeehouses…"

…

"The Taj Mahal is a marble mausoleum in Agra, India."

"Saint Peter's Basilica is a church located in Vatican City."

…

WIKIPEDIA

Who are customers?
Who is selling ice cream?

Problem

A queue of customers patiently waiting to buy ice cream

"Häagen-Dazs is an American ice cream brand …"

"Starbucks Corporation is a chain of coffeehouses…"

…

"The Taj Mahal is a marble mausoleum in Agra, India."

"Saint Peter's Basilica is a church located in Vatican City."

…

WIKIPEDIA

We introduce COFAR

A dataset for Image Search requiring COmmonsense and FActual Reasoning

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

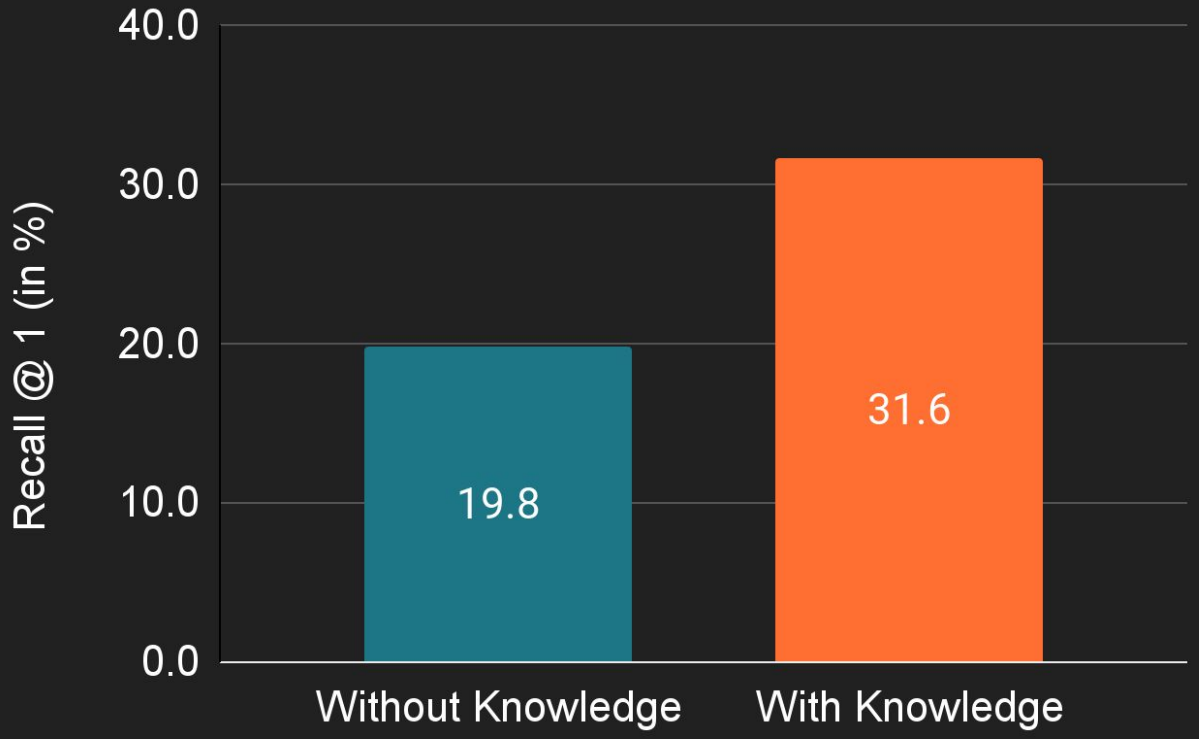
“Two people getting married in front of a tower in Paris”

“Two people showing an interesting to purchase a watch”

Dataset Statistics

- 25K+ Images
- 40K+ Text queries
- 5K+ Named Entities
- Includes banks, restaurants, actors, athletes, landmarks, monuments, and more

Our Baseline Model (KRAMT)

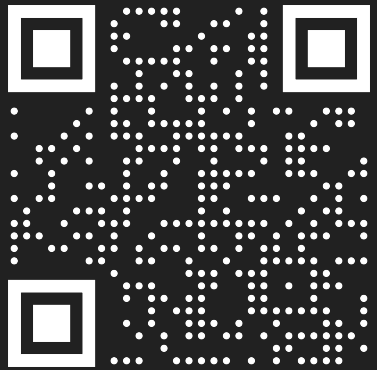


AACL IJCNLP 2022

Thank you!

Visit our project page for Code and Dataset

https://vl2g.github.io/projects/cofar



COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

accenture