

Caption generation for Images in News articles

Problem Statement

Informal Description: Given an image and an associated news article, generate a caption that will be associated with the image displayed in the news article with the given text.

Semi-Formal Description: Given an image I with one or more channels and a article with n tokens (x_1, x_2, \dots, x_n) , generate a caption with m tokens (y_1, y_2, \dots, y_m) where $m \ll n$

News Image captioning with pre-trained models

- [Source:https://ieeexplore.ieee.org/document/10219679](https://ieeexplore.ieee.org/document/10219679)
- Knowledge Prompt-Driven method
- Key modules: **commonsense knowledge reasoning** and **multimodal context reasoning**.
- Commonsense knowledge reasoning aims to query the commonsense knowledge of named entities that induces a pre-trained language model for commonsense reasoning using context-aware knowledge prompts.
- Multi-modal context reasoning aims to model multimodal contexts (news articles, images, and commonsense knowledge) for news caption generation.
- Cross-model consistency objective is used to maintain the consistency between news caption and the news image.
- Datasets: GoodNews and NYTimes

Challenges

- Rate limiting from NYT APIs causing data collection problems
- Training of models taking a long time even for a single epoch
- Limitations around GPUs
- For pre-trained models: challenging to train (fine tune models) them on NYT dataset
- Measuring the correctness of the caption - BLEU & METEOR fit better for Machine translation
- Choosing lengths of articles & VisualGenome information

Overview of Dataset

- Extracted 50k keys from the NYT dataset.

caption for the key: 5397a3d038f0d80b34dcbdd6
: {'0': 'The United States team training in São Paulo.'}

article for the key: 5397a3d038f0d80b34dcbdd6
: RONALDO RETURNS Cristiano Ronaldo gave Portugal a major confidence boost ahead of the World Cup when he made a successful return from injury. Playing his first match since the Champions League final on May 24, Ronaldo helped reignite Portugal's stuttering attack in a 5-1 romp over Ireland at MetLife Stadium in East Rutherford, N. J.

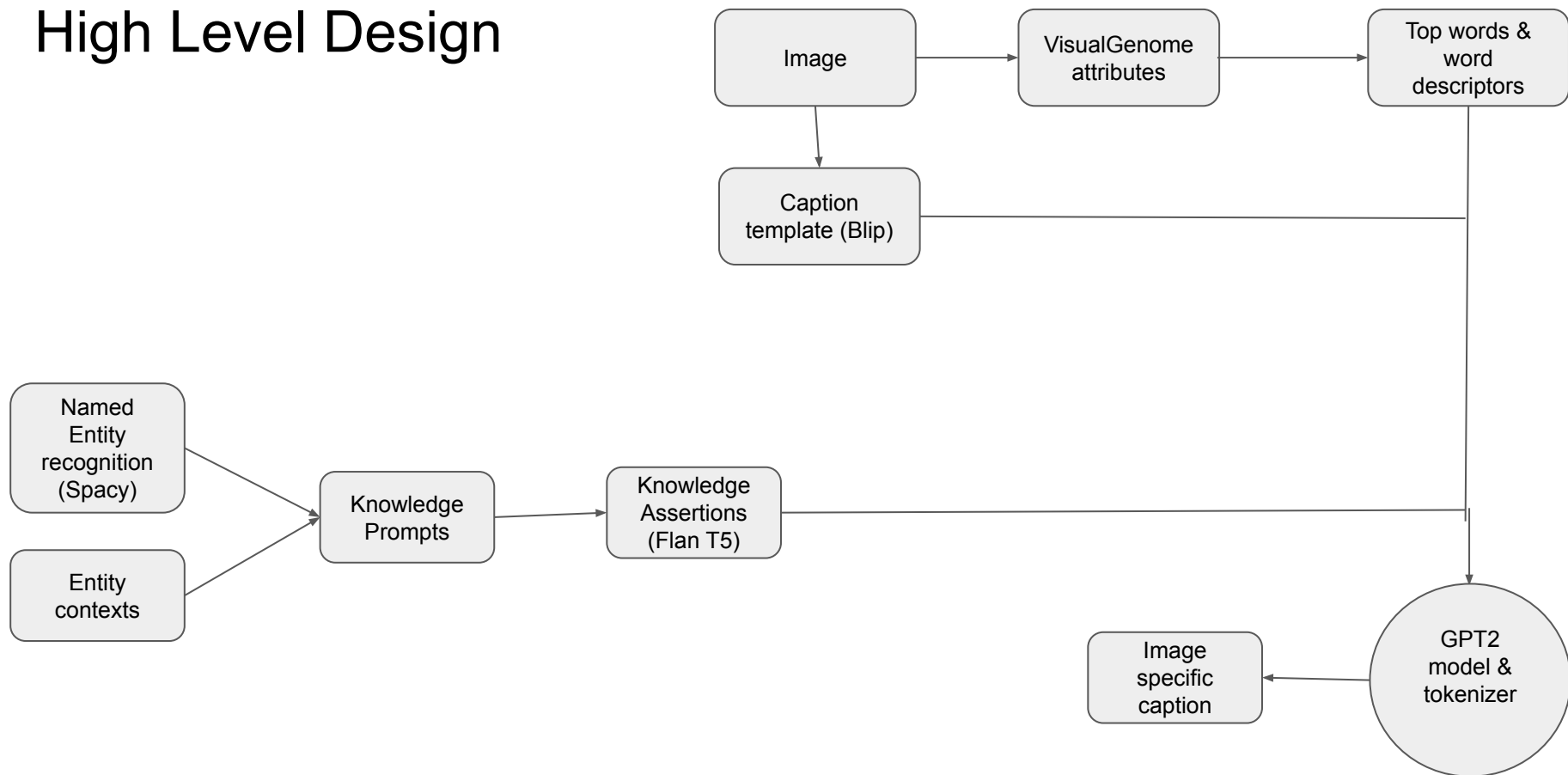
Ronaldo, the world player of the year, did not score and left after 64 minutes but showed glimpses of his best and came through the friendly unscathed after missing his country's last two warm-up matches because of problems with his left knee and right thigh. (REUTERS)

0



5397a3d038f0d80b34dcbdd6_0.jpg

High Level Design



Steps we followed

1. Get **Named entity** and **entity contexts** for Article.

```
('Portugal', 'Cristiano Ronaldo gave Portugal a major confidence')  
( 'the World Cup', 'boost ahead of the World Cup when he made')  
( 'first', '. Playing his first match since the')  
( 'the Champions League', 'first match since the Champions League final on May')  
( 'May 24', 'League final on May 24, Ronaldo helped')
```

2. Get **Knowledge prompt**

Knowledge Prompt:

Generate a summary based on the named entity: Portugal based on the news context: Cristiano Ronaldo gave Portugal a major confidence

Generate a summary based on the named entity: the World Cup based on the news context: boost ahead of the World Cup when he made

Steps we followed

3. Feed Knowledge prompt to get Knowledge assertions

Knowledge Assertion: ['cristiano ronaldo gave portugal a major confidence boost on tuesday.']

Knowledge Assertion: ['the san francisco 49ers beat the san francisco 49ers in a 5-1 romp, and the san francisco 49']

4. Get attributes and **generic caption** from the image

crouch, Probability: 0.297

running forward, Probability: 0.090

athletic style, Probability: 0.072

Generic caption: a group of people playing soccer on a field

5. Feed entity context, knowledge assertions generic caption in GPT to get caption

Results

0

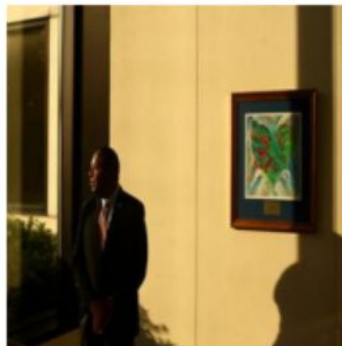


Actual caption:The United States team training in São Paulo.

Generated Caption 1: The U.S. team trains in São Paulo, getting ready for the World Cup. (BLEU: 0.44)

Generated Caption 2: Cristiano Ronaldo trains with the Portugal team for the World Cup. (BLEU: 0.35)

0



Actual caption: This photograph of Leroy Smith aiding an ailing man during a rally in Columbia, S.C., on July 18 was posted on Twitter

Generated Caption 1: Leroy Smith, wearing his uniform, helps an unwell man at a rally in Columbia on July 18. (BLEU: 0.45)

Generated Caption 2: Leroy Smith helps a man during a rally in Columbia. (BLEU: 0.23)

Thank you.