Learning Through Interaction:
A Feedback-Driven Image Captioning System

Enhancing Image Understanding through User Interaction

Rohitha Ravindra Myla , Aishwarya Singh

Motivation

User Specific Images with less annotated training data

How do we fix this?

Maybe ***you*** can fix this

We propose an IC model with Adaptive Attention integrated with interactive user-feedback

Why is there a need for more enhanced models?

- 1. Assistive Technologies:

  Helping visually impaired individuals understand their surroundings by describing scenes with user inputs like “Explain objects in the room.”

2. Content Creation:

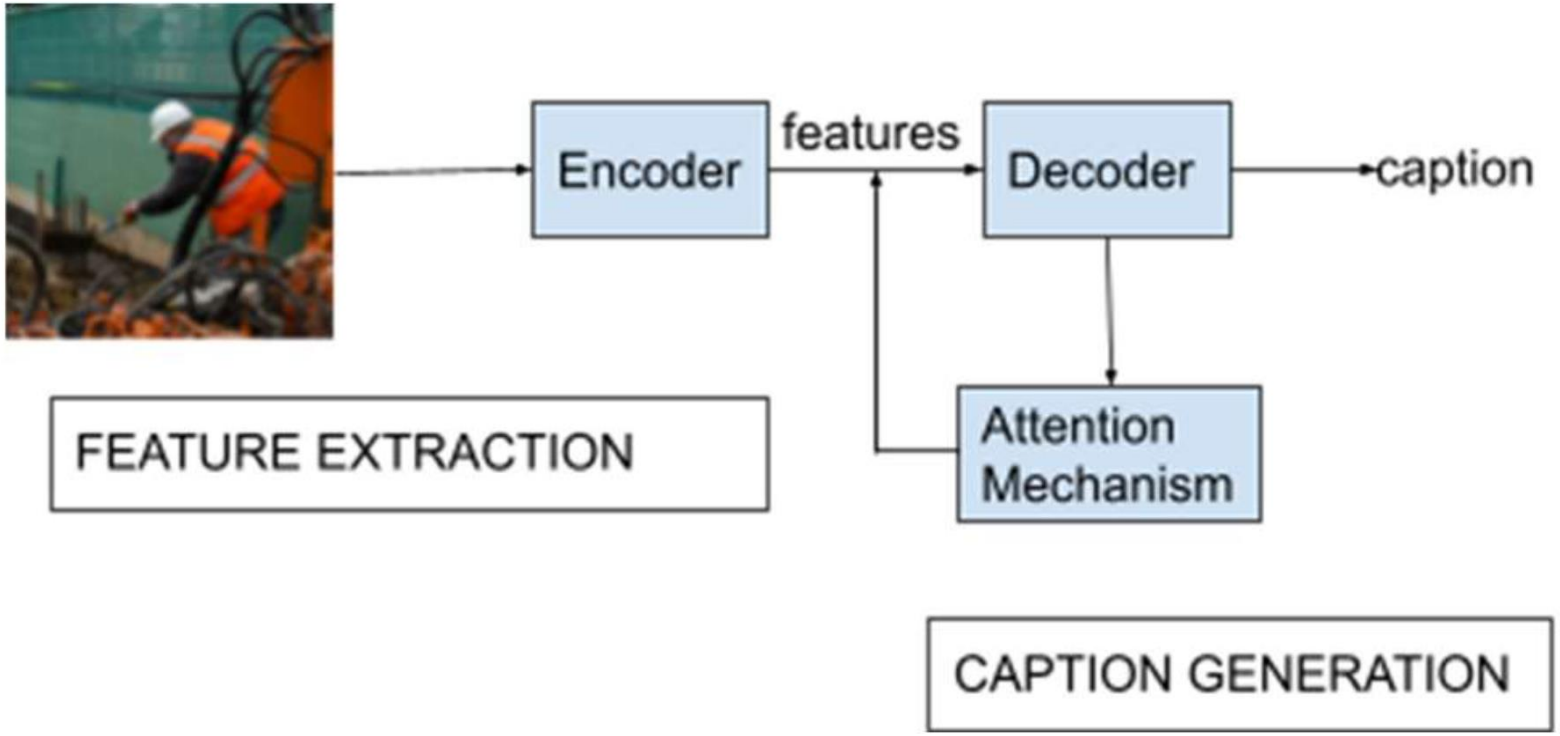
  Social media caption generation based on user prompts like “Make it funny.”

Our Approach and Intuition

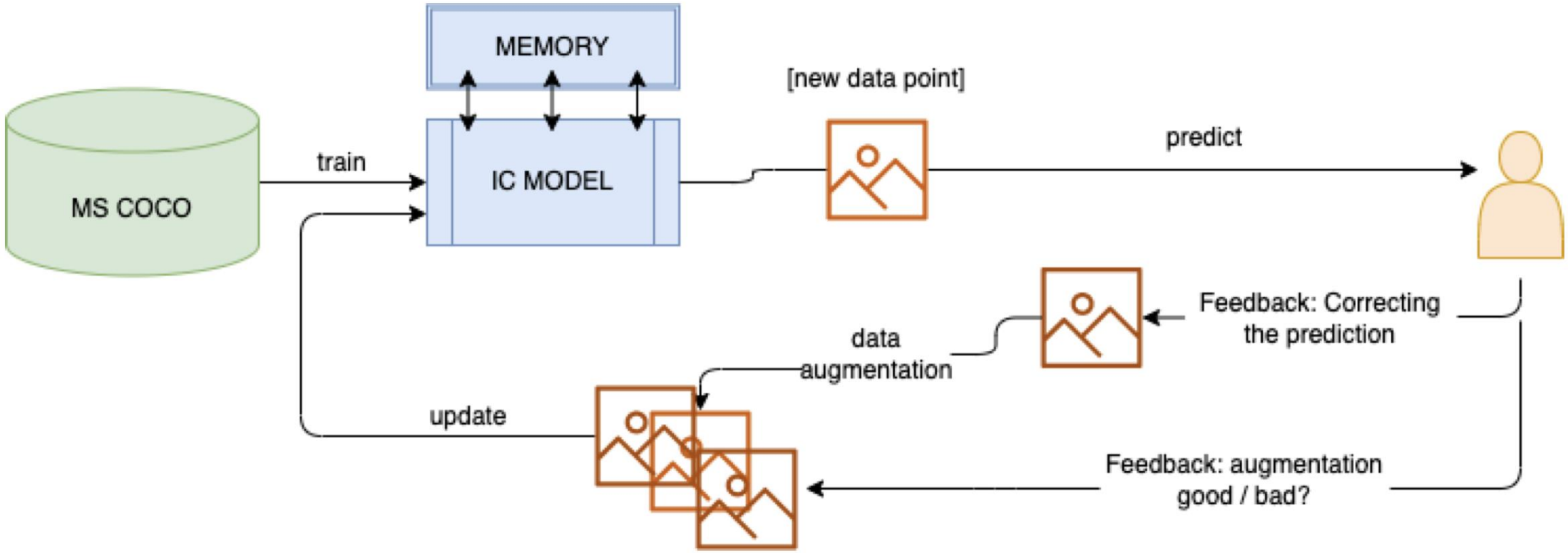
- Implementing Adaptive Attention Mechanism
- Integrating a user feedback framework

General Architecture

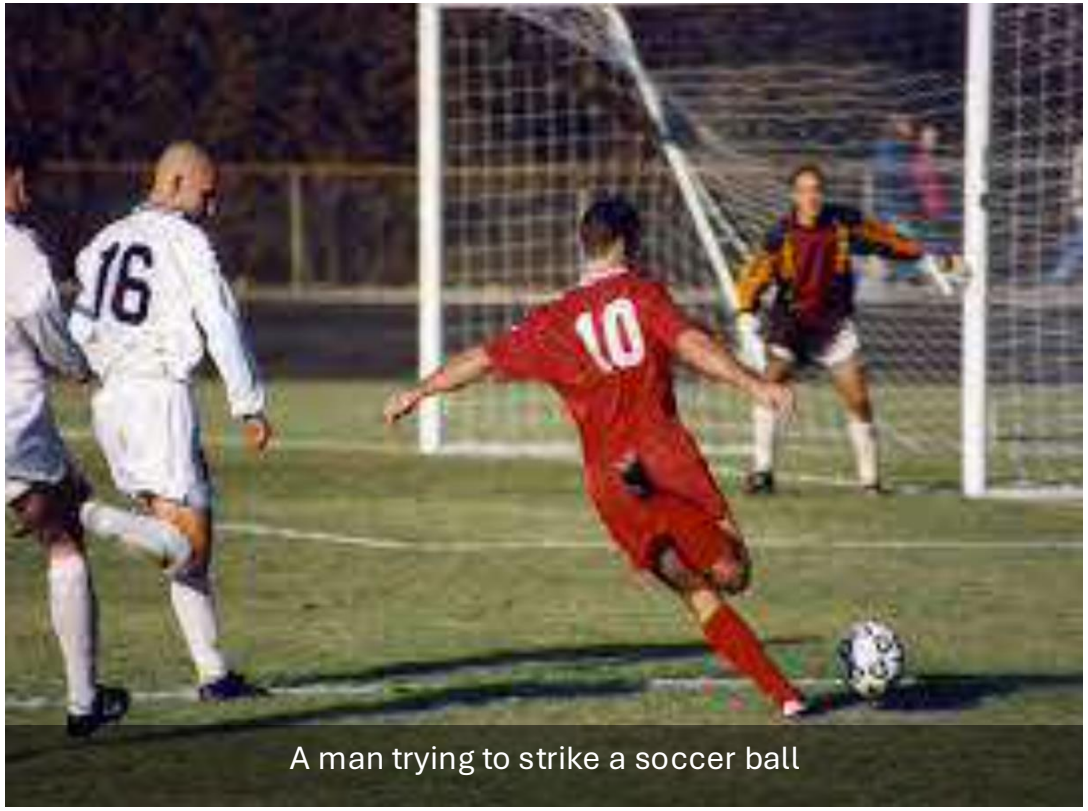
CNN encoder and RNN decoder

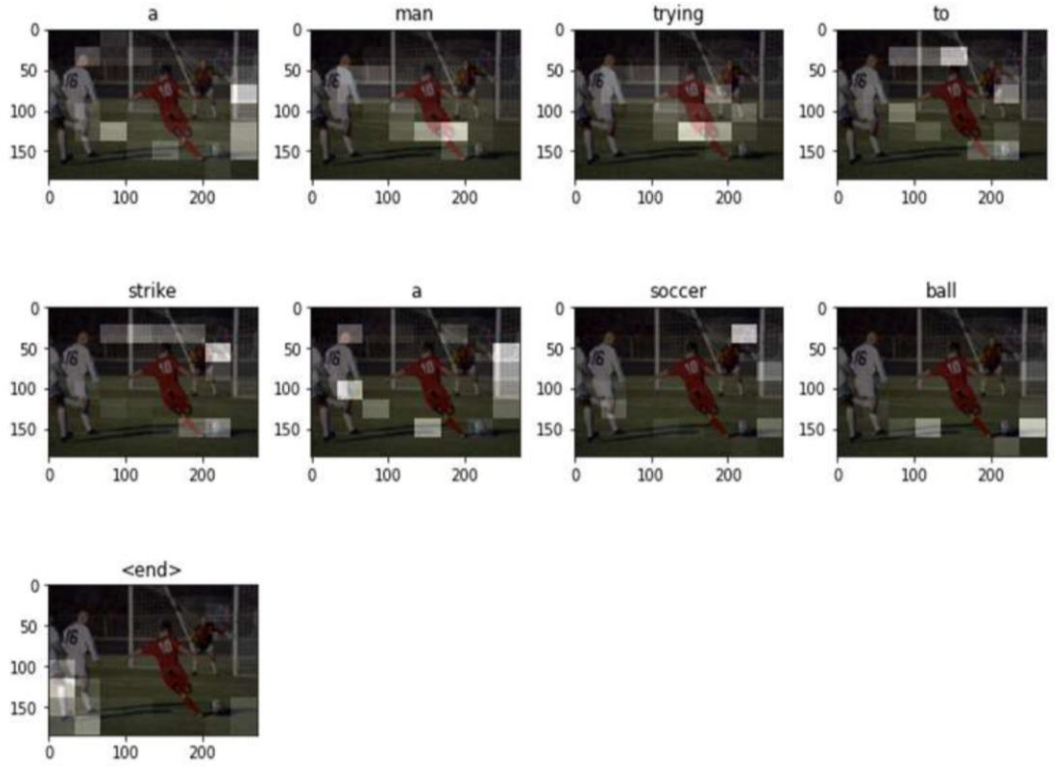


How does the user feedback work?

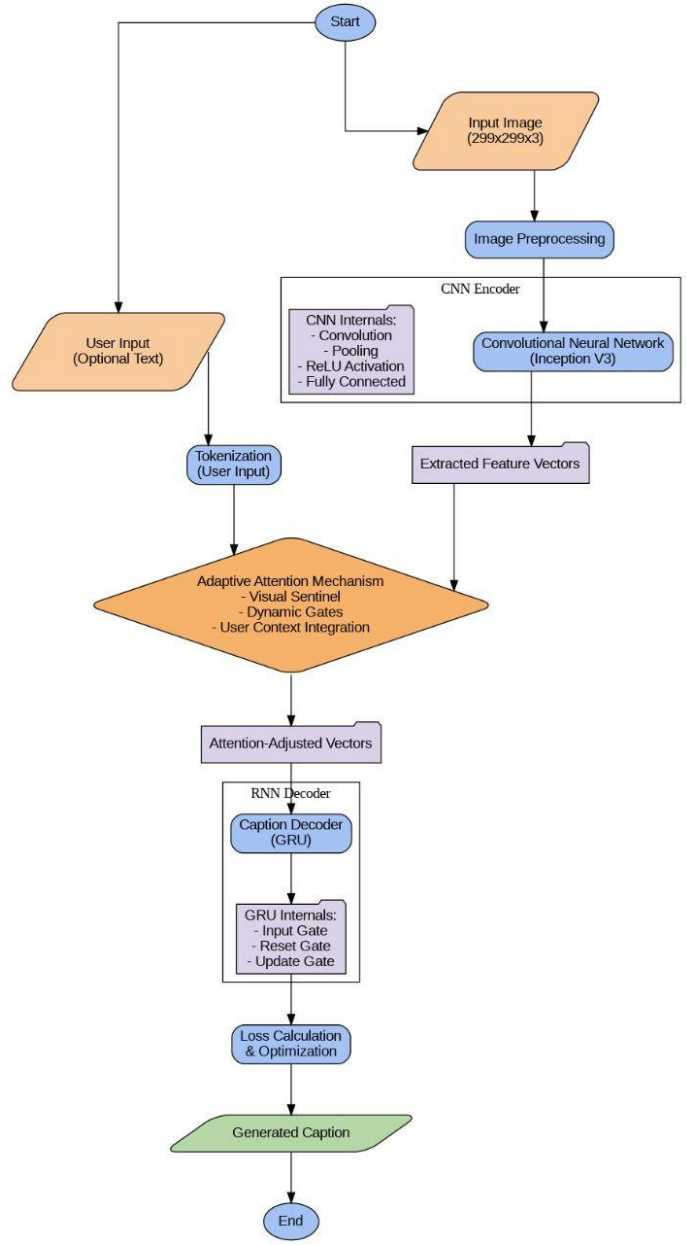


Why Adaptive Attention?



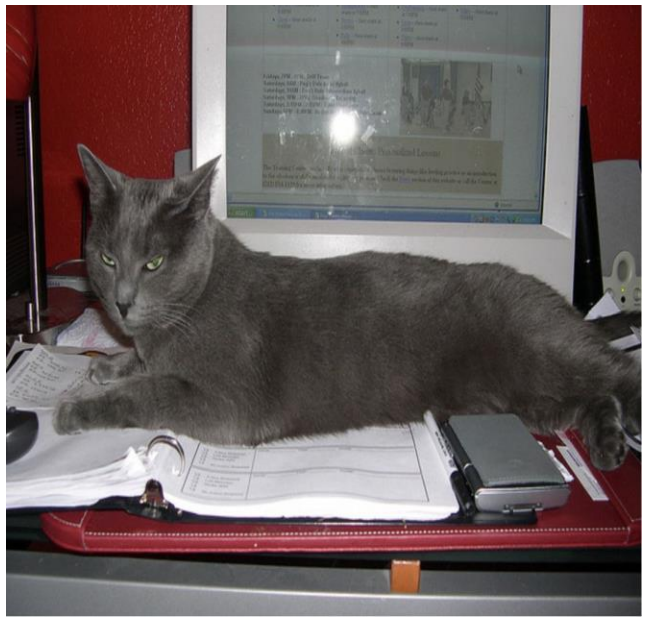


Architecture Deep Dive



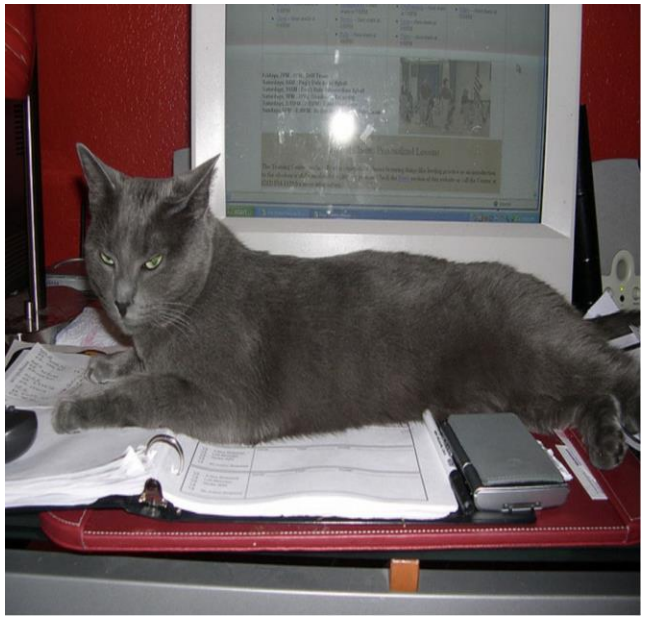
Achievements

Without user input



A grey cat lying in front of a computer monitor

With user input



An angry grey cat lying in front of a computer screen

Achievements

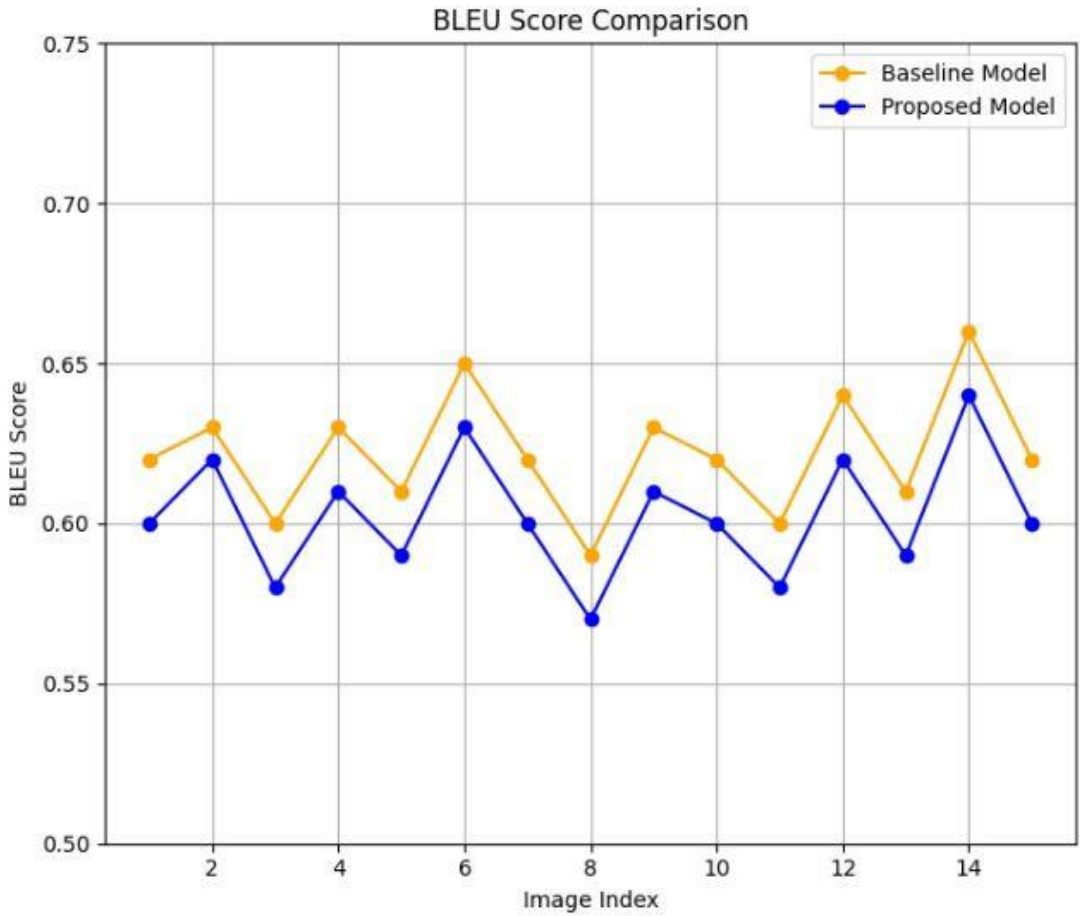
Without user input

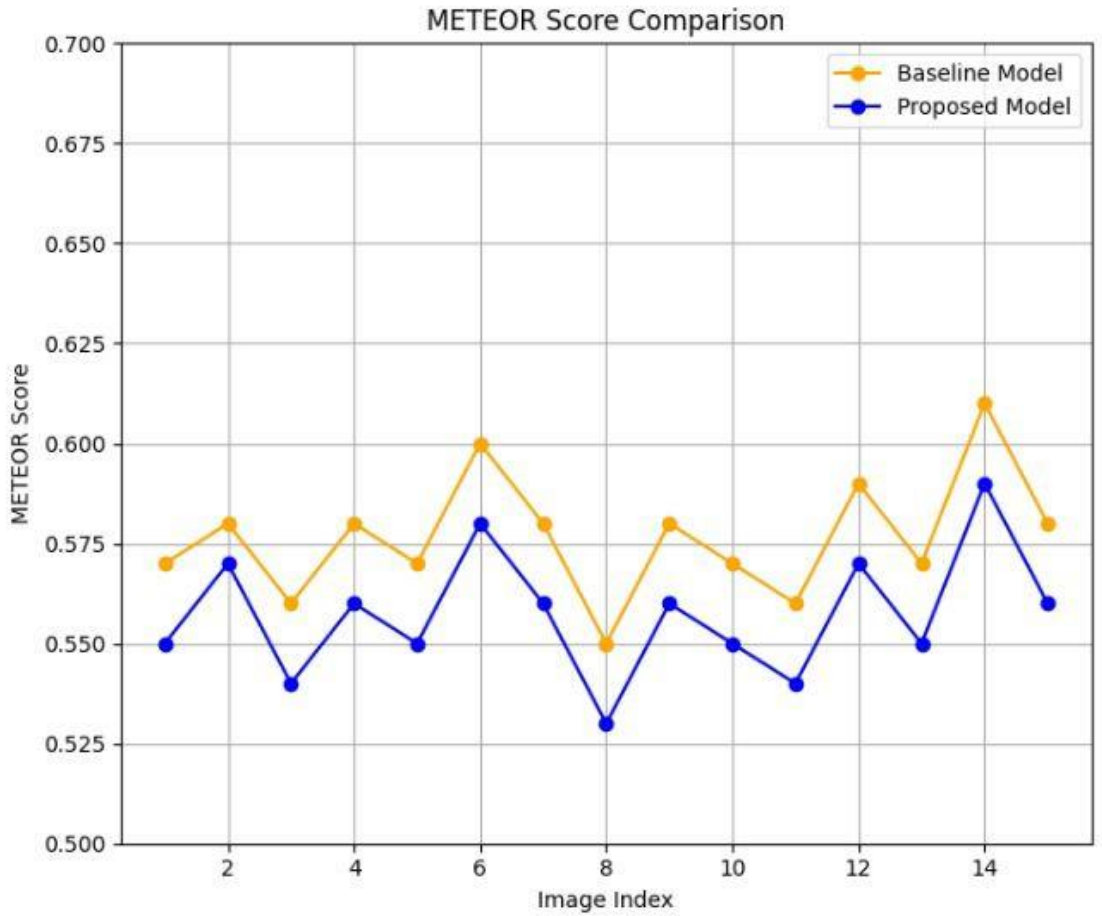
A car parked on the side of the road

With user input

A Ford car parked near a quiet street, illuminated by streetlights

Learnings


BLEU Score Comparison
Baseline Model
Proposed Model
BLEU Score
Image Index
0.75
0.70
0.65
0.60
0.55
0.50
2
4
6
8
10
12
14


METEOR Score Comparison
Baseline Model
Proposed Model
METEOR Score
Image Index
0.700
0.675
0.650
0.625
0.600
0.575
0.550
0.525
0.500
2
4
6
8
10
12
14

Challenges

- Large datasets like COCO do not always represent user-specific contexts or niche domains.
- Human feedback, while valuable, is limited in volume and variability.

Solution?

Data augmentation of user feedback

Further research

- Incorporating active learning techniques to select the most informative samples for memory replay.
- Extending the approach to support multilingual image captioning, beneficial in scenarios with limited annotated data.

Questions?

Thank you!