

ImagiNarrate

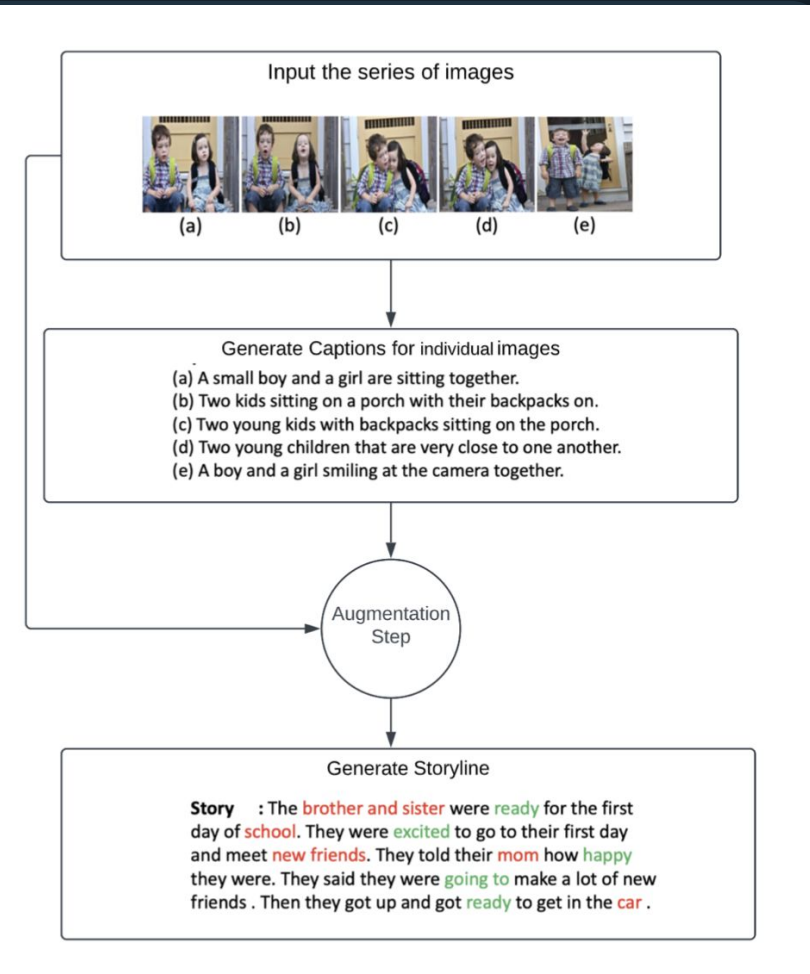
Building a Narrative with Images and Generated Captions

Team:Group 46

Asmita Chotani, Apurva Gupta, Chetan Chaku,
Priya Nayak, Rutuja Oza

Motivation

- Growing interest in visual captioning and limited ability of machine to understand complex human-like scenarios.
- **Baseline**: AREL model - generates story based on sequence of images.
- **Proposed model**: Generates caption and augments with image sequence to produce story.
- **Future scope**: Creating novels from comic books for the visually impaired audience.

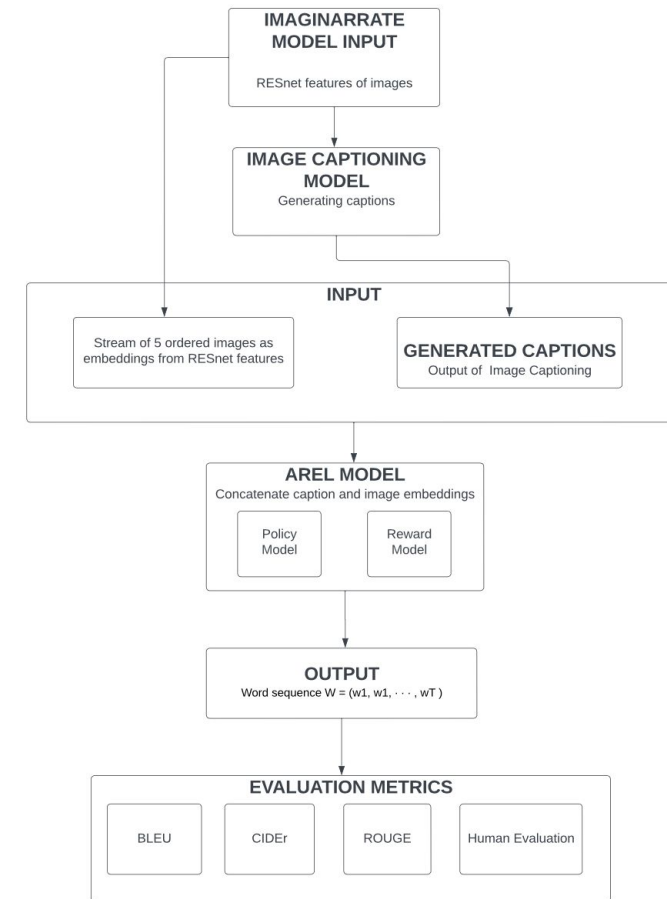


Problem Definition

- **Input:** Sequence of 5 ResNet-152 features
- **Output:** A story related to the sequence
- **Training:**
 - Generation of caption per image
 - Augmentation of ResNet-152 embeddings with caption embeddings
 - Generation of story
- Incorporating captions along with images provides informative descriptions for the model to understand & build cohesive stories.

Proposed Solution

1. **Caption generation model:** LSTM based encoder - decoder structure
2. **Story generation model:** RL framework. Encoder uses GRUs & decoder uses GRU-RNNs.
3. **Dataset:**
 - a. ResNet-152 features of VIST images
 - b. Ground truth stories from VIST dataset
4. **Hyperparameters:**
 - a. **Vocab size:** 9839 words.
 - b. **Epochs:** 60 - computational constraints
 - c. 210819 unique images



Result

- **Natural Language Generation Metrics for Evaluation:**
BLEU, CIDEr, ROUGE for generation fluency.
- Increasing training epochs could **improve generation fluency**.


Model	Bleu_1	Bleu_2	Bleu_3	Bleu_4	ROUGE	CIDEr
AREL Model - 100 epochs	0.642	0.391	0.231	0.140	0.295	0.097
ImagiNarrate Model- 60 epochs	0.582	0.360	0.211	0.126	0.298	0.045

Result

- **Human Evaluation:**
Human feedback via circulated Google Form.
- **Questions Asked:**
 - How satisfied were you with the overall story generated for the given image?
 - How engaging was the story?
 - How easy was it to understand the story?
 - How well did the story flow and maintain a coherent narrative?
 - How well did the story integrate with the given image?

ImagiNarrate Story Output 1

For the given set of images and story generated. Please answer the below 5 questions.



STORY GENERATED: " The family gathered for the family dinner . Everyone was having a great time . They had a lot of food . She was so happy to see the kids . After the night , we all had a great time ."

How satisfied were you with the overall story generated for the given image?

	1	2	3	4	5	
Strongly Disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly Agree

How engaging was the story?

	1	2	3	4	5	
Not at All	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very