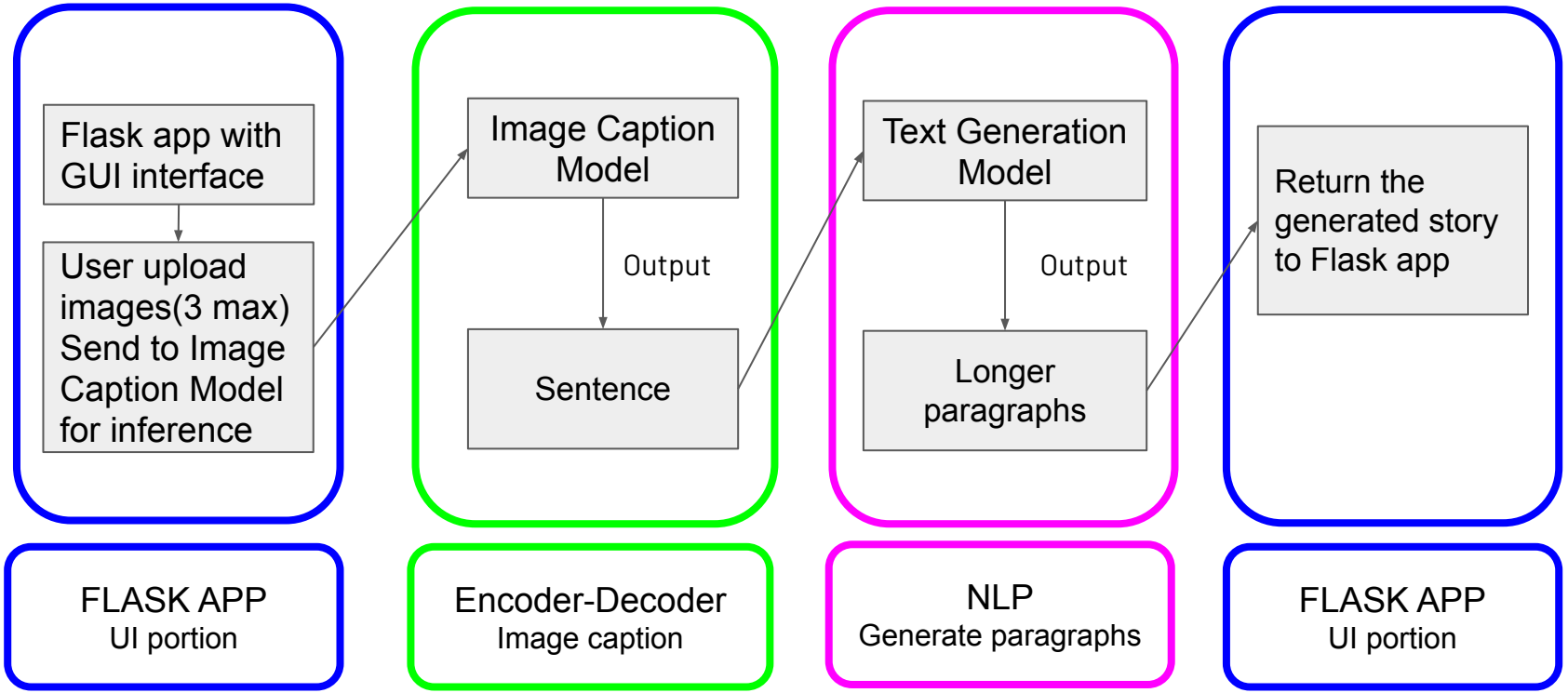
Image Story Generator

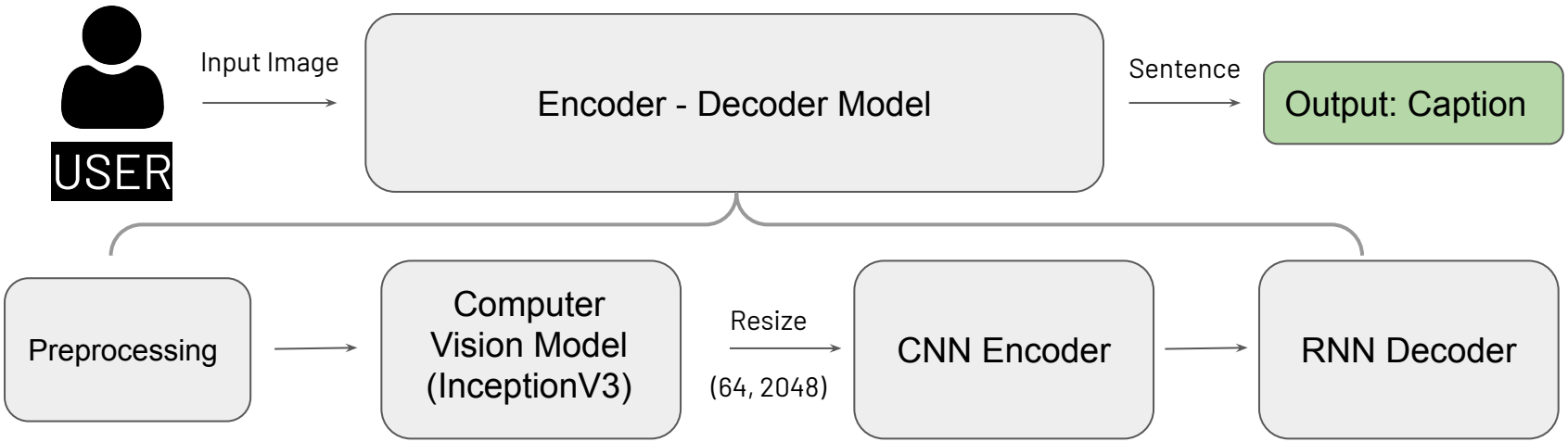
Chang Xuan, Guan Kiong, Josephine, Rebecca, Shaun

Project Flow



Models

Image to Caption



Inception V3's expected format:

- Normalize the image pixels to the range of -1 to 1
- Resize images to (299,299)

- Pre-trained on Imagenet
- Output layer: last convolutional layer of Inception V3
- Output: Convoluted Features (8, 8, 2048)

- Single fully connected layer (256 nodes)

- GRU
- Bahdanau Attention (used to give weightage to keywords)

Dataset: MS-COCO

- over 82,000 images
- caption annotations for each image

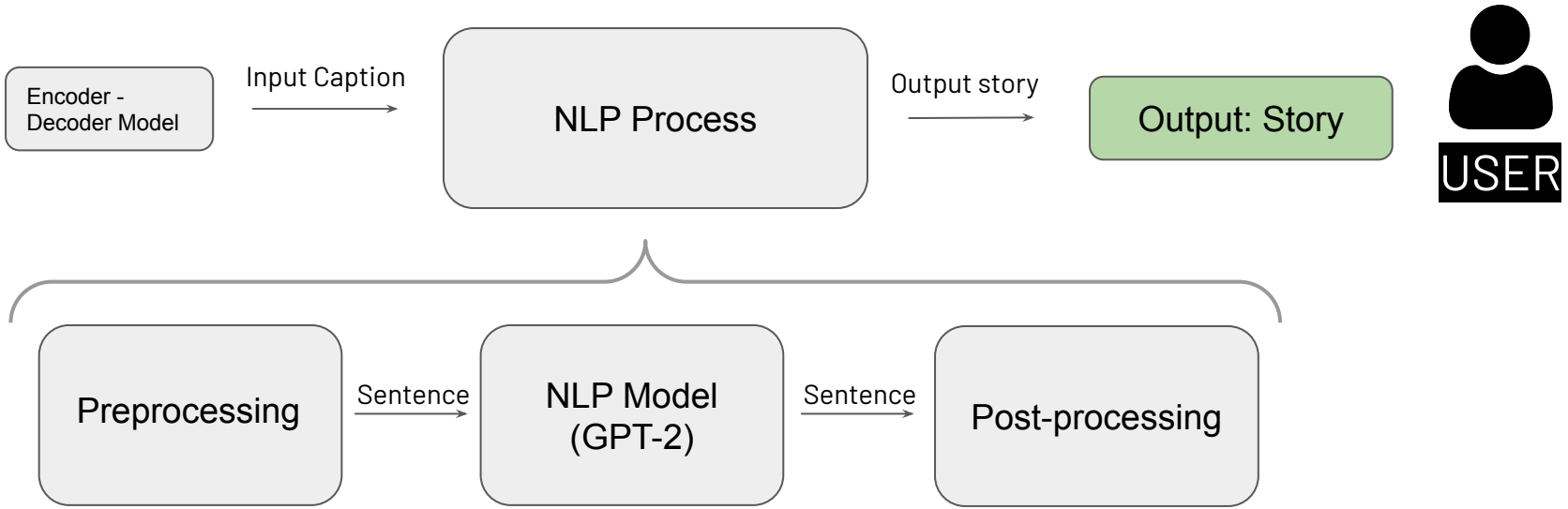
- Examples:

'<start> a person soaring through the air on skis <end>'

'<start> A Do Not Enter sign is posted along a road leading to a stadium. <end>'

'<start> Small child in a high chair eating off of a green plate. <end>'

Caption to Story



- Cleaning (removing <> etc.)
- Convert past tense
- Grammar correction
- Add narrative hook

- Remove incomplete sentences

Dataset: Narrative Hooks

- We obtained 9,000+ narrative opening lines compiled from a crowd-sourced dataset (Janelle Shane's Novel First Lines Dataset) of input from users and actual novels

- Examples:

    - It was love at first sight.
    - A secret is a strange thing.
    - "I don't understand."
    - All this happened, more or less.
    - It was the year of electrocution.
    - I lead a double life.
    - Sally Louisa Tomkins stood her ground.
    - "Watch out!" yelled Pete Crenshaw.

- These Narrative Hooks are used as the opening line(s) and combined with the Caption — before it is passed into the NLP Model to produce the Story

- Strong opening lines help

    - introduce some novelty — a good balance against the relatively descriptive, sterile Caption
    - creates synergy with GPT-2
    - pique readers' curiosity
    - create an emotional investment or connection
    - provide entertainment, via humour, suspense, or shock

Demo

Area of Improvements & Stretch Goals

- **Areas of Improvement**
    - Pick a Narrative Hook that is thematically similar to Caption
        - Cosine similarity
        - Jaccard index
    - Speeding up model inferences
- **Stretch Goal #1: Story from multiple images**
    - Current - each Story is built on 1 input image
    - Future - each Story is built on multiple input image
- **Stretch Goal #2: Develop context-specific stories**
    - Based on user segment - e.g children-friendly stories & fairy tales
    - Based on genres - e.g romance, mystery, thriller, sci-fi

Image captioning

- 20k Instagram posts with captions from celebrities: https://www.kaggle.com/prithvijaunjale/instagram-images-with-captions
- Flikr 30k https://www.kaggle.com/hsankesara/flickr-image-dataset

https://gombru.github.io/2018/08/01/InstaCities1M/

Project / Github	CV model	NLP model	Dataset	URL
SemStyle	Inception-v3	GRU + term/word embedding vectors	Descriptive image captions - MSCOCO Styled text - romance novels corpus	https://arxiv.org/abs/1805.07030
StyleNet	CNN	LSTM	FlickrStyle10k (self-built based on Flikr 30k dataset)	Paper: https://www.microsoft.com/en-us/research/uploads/prod/2017/06/Generating-Attractive-Visual-Captions-with-Styles.pdf Git: https://github.com/kacky24/stylenet
Attend to You: Personalized Image Captioning with Context Sequence				https://github.com/cesc-park/attend2u

Project / Github	CV model	NLP model	Dataset	URL

Tasks

- Flask, Web App (Make it flexible to either allow 1 or up to 3 images) - **Guan Kiong**
  - Interface with api from CV portion and NLP portion.
  - Either to display Both CV and NLP portion or only display the end outcome(Only display story)

- TensorFlow Image Captioning Model - **Rebecca, Chang Xuan**
  https://www.tensorflow.org/tutorials/text/image_captioning
  - Input: 1 Image in jpg/png
  - Output: 1 Sentence
  - Task Done: Verified model works, pushed code

- GPT-2 Model - **Josephine, Shaun**
  https://www.analyticsvidhya.com/blog/2019/07/openai-gpt2-text-generator-python/
  - Input: Sentence, Output: Paragraph
  - Task Remaining:
    - Input gpt2.py

- ??? - Re-convene for Data Pipeline

Proposed Sequence

Input Image

Computer Vision Model

Hashtags

NLP Model

Sentences

Sentences + Hashtags

Output: Caption

Pick from our #hashtags

Multi-label object classification

- Objects detected are converted to #hashtags (e.g #nature #cars)

Sentence Generator

- Model generates sentences using #hastags obtained from images

Pre-trained #hashtags

Our Pre-Trained HuggingCaptions Model

Sentences

HuggingCaptions model

- We pre-train a few GPT-2 models on selected, popular #hashtags
- Model generate sentences using popular #hashtags