Assignment #5

Image Captioning

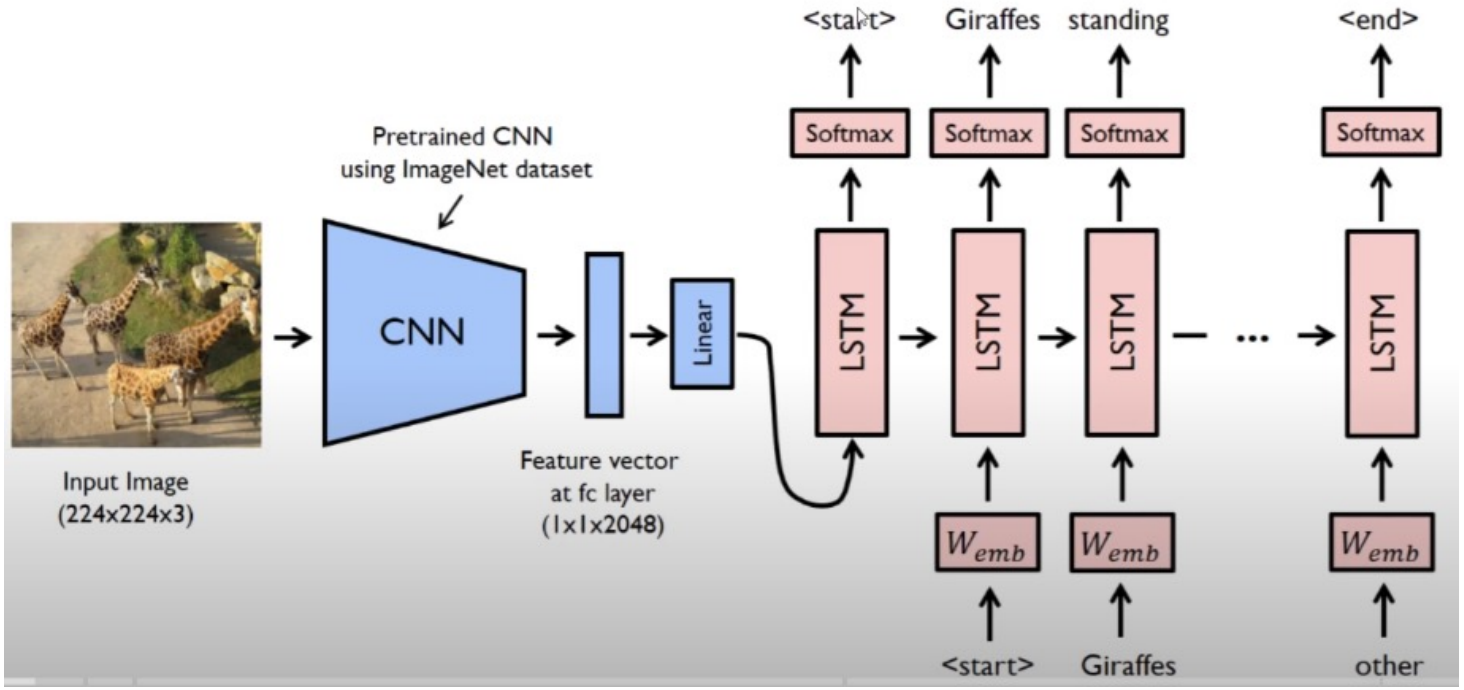
Due on Dec 21, 11:59 pm

Overview

- In the previous assignments, we implemented multiple image classification tasks.
- In this assignment, you will design and train a neural network which combines CNN and RNN to process an input image, then output a sequence that describe the image.
- You are free to use pre-trained models like ResNet or LSTM as your backbone structure.

Image Captioning

Image captioning is an interdisciplinary research problem that stands between computer vision and natural language processing.

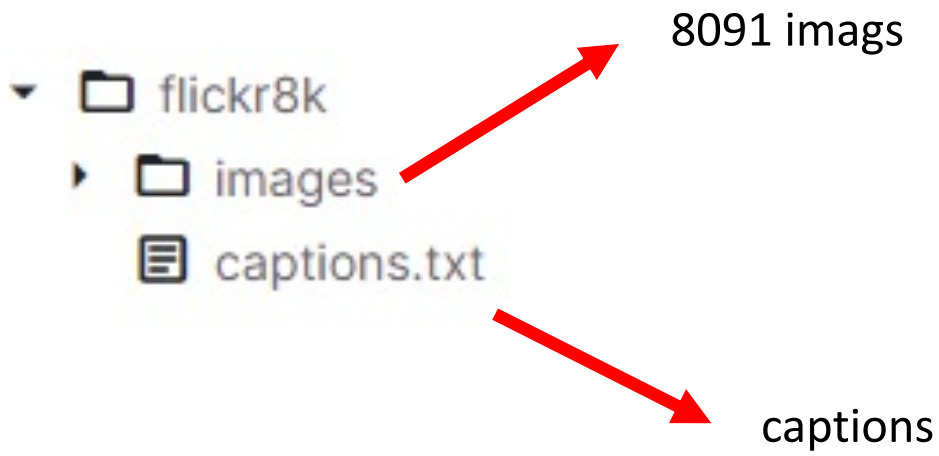


Flickr8k Dataset

- Flickr8k-Images-Captions
- Collected by Alexander Mamaev.
- Sentence-based image description and search
- Consisting of 8,091 images that are each paired with five different captions

A child in a pink dress is climbing up a set of stairs in an entry way .

Assignment #5 Dataset



Your task

- We have code skeleton for you guys.
- https://colab.research.google.com/drive/1E96yjndJyBTAEEcgSth-RyAVqd4H1WcW?usp=sharing
- Design a convolutional neural network to do image captioning.
- The images provided are of different resolutions. You'll need to resize the images into a fixed size of your own choice.
- To get a high accuracy, you'll need to experiment with different filter sizes, different number of layers, and other design principles discussed in class to figure out a network architecture that works best.
- You'll also need to try data augmentation, dropout, batch normalization as well as different optimizers and other tricks to boost performance.

Things you cannot do

- You cannot copy trained models from others.
- You cannot copy a whole page of code from the Internet.

Any violation will result in no points!

Submission

Submit your code + report**(pdf file)** to the CU.

- The report include:
  - **Additional works**.
  - What you have experimented for performance improvements. This could include experiments that work and those that do not work.
  - Other findings if any
- Grading on report will mostly depend on the report quality. Please show your effort.

Grading

- Code: 85 points (A)
- Report: 15 points report