

ECE/CS 6524 – Deep Learning

Project Assignment #3: Image Captioning with RNNs

In this assignment, you are asked to implement and train recurrent neural networks (RNNs) (using Keras) for image captioning. For this assignment, you may need to use GPUs to speed up the training of RNNs. If you have access to GPUs locally or remotely (such as VT's ARC cluster), you are ready to follow the tutorial below to train RNNs. Otherwise, use [Google's Colaboratory](https://colab.research.google.com/) (Colab) for your access to free GPUs.

1. Please read a tutorial on using Keras to train an RNN for image captioning: **"Image Captioning with Keras"** (<https://towardsdatascience.com/image-captioning-with-keras-teaching-computers-to-describe-pictures-c88a46a311b8>). Specifically, you need to understand some basics on processing texts and images using Keras.

2. Implement and train a LSTM to caption the images in the Flickr 8k dataset, which can be downloaded from the Canvas system.

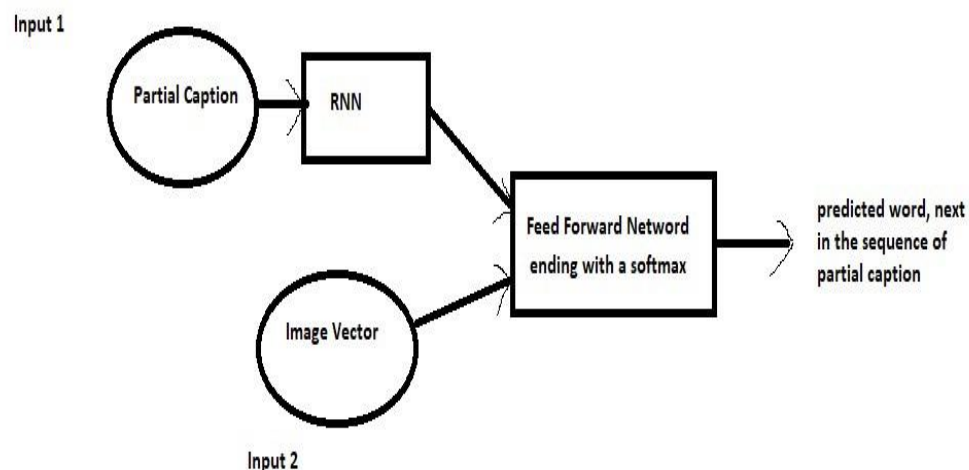


Fig. 1. A Block Diagram of the RNN approach to image captioning.

A conceptual diagram of the image captioning approach is illustrated in Fig. 1, where image features are extracted by pretrained CNNs such as VGG/ResNet/Inception. The recurrent neural network (RNN) can be implemented with LSTM or GRU network.

Train an RNN with the training dataset (6000 images) and test it with the test dataset (1000 images). Evaluate the performance of image captioning with BLEU Score (<https://www.digitalocean.com/community/tutorials/bleu-score-in-python>).

You need to prepare **a written report in the ICCV format** (ICCV template/instruction docs attached), including the following sections: (1) Introduction/Problem Statement; (2) Approach(es); (3) Experimental Results; (4) Discussion and (5) References (if any).

In addition, you need to attach your **implementation codes** as separate files to the report.