**Documentation:**

The image caption model is an end to end model, that takes the input(text caption and image, filename) to train the model and generate caption on the test set.

**Experiment:**

We experiment with several variants of the model:

* **RNN-rand** Our baseline model where all words are randomly

initialized and then modified during training.

* **RNN-static** A model with pre-trained vectors from Glove.

All words including the unknown ones that are randomly

Initialized at are kept static and only the other parameters

of the model are learned.

* **RNN-non-static** Same as above but the pre-trained vectors

are fine-tuned for the task.

**Dataset:**

1. Used Fickr8k dataset, having text data and image data separately. The text data are in .csv file mapped with the file name of the image.

Link to Dataset: <https://drive.google.com/drive/folders/1N66PEs1VmwZCkODgf6la5Rj2cSi34rBm?usp=sharing>

each image has 5 captions. For the basic implementation, we used 1 caption. For late part of the experiment, we will use more captions. The images are in jpg format.

2. from the dataset, we created 'dataframe.pickle' file to arrange the data information to use in code, included in

'code' folder.

3. The path to the dataset and other resource files are given in code (e.g. dir\_Flickr\_jpg,dir\_Flickr\_text, dir\_Flickr\_res), it has to be changed accordingly.

**Structure of the code:**

1. Though it is an end to end model, for debugging purpose and tracking time, we have implemented it incrementally.

2. Saved the intermediate results (e.g. train, validation, test data set, trained model and so on) in files which are in the ‘codebase’ folder

3. The final stage of the code that generate caption on test set and evaluate the model, uses those files.

**How to Run:**

1. Run command:

python image\_caption.py

2. It will execute the the methods in main() method sequentially to generate end result

**Results:**

1. The final portion of the code generate results saved in file ’bleu\_score’, that is in 'result' folder. That has the final result for the variants of the models of the task.

2. There are some intermediate result files on the image and text data, that are also included in codebase folder.

3. Results of different variant of the model is as follows: