ITE4052 Computer Vision (Spring 2024)

Programming Assignment 3

- Deadline: June 13th 11:59 PM.
- Submit the all codes (.zip) and a report summarizing the results (pdf or docx) on the LMS.
- You can use either English or Korean for the report.
- Late submission will get the half score.
- If you use ChatGPT, you can be caught for plagiarism (immediate F). Don't use it.
- It will be helpful to read the commented part carefully.
- Do not use external knowledge (e.g. external dataset, pre-trained model).

Description of the data

Dataset

- COCO dataset 2014 (for 1,2,3)
  - Training image: 80,000
  - Validation image: 40,000

Image feature

- Raw images take up nearly 20GB of space so we have not included then in the downloads.
- We have preprocessed the data and extracted features for you already.
- For all images, we have extracted features from the fc7 layer of the VGG-16 network pre-trained on ImageNet, and these features are stored in the files **coco_captioning/train2014_vgg16_fc7.h5** and **val2014_vgg16_fc7.h5**.
- We have stored the URLs of the training and validation images in the files **train2014_urls.txt** and **val2014_urls.txt**.
  → This allows you to download images on-the-fly for visualization.

Caption

- Each word is assigned an integer ID, allowing us to represent a caption by a sequence of integers.
- The mapping between integer IDs and words is in the file **coco_captioning/coco2014_vocab.json**, and you can use the function **decode_captions** from the file **CV/coco_utils.py** to convert NumPy arrays of integer IDs back into strings.

Test Data (only for 4.Inference on NICE Challenge Dataset)

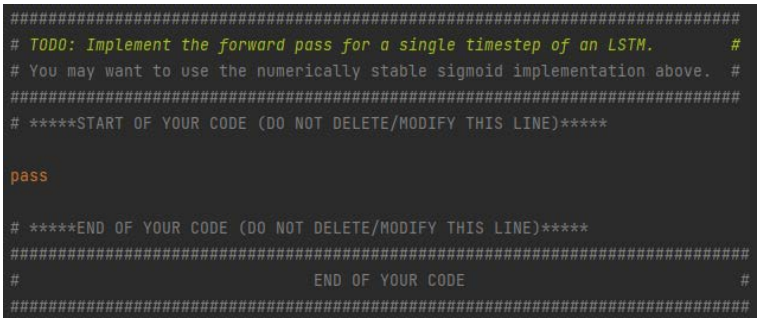
- You can download the test data in here(size: 200MB) Test data for this assignment is sub data of Nice challenge 2024. There are 1000 images in test data. After test your code with these images, please make the result as json and upload your **.json (The extension must be ".json")** to the website. The descriptions for method to submit the results are on the **website.**

  Website address: http://mmai.hanyang.ac.kr:81/#/

- ✓ You can load all of the COCO data (captions, features, URLs, and vocabulary) using the **load_coco_data** function from the file **CV/coco_utils.py**. If you get an error (TypeError), the URL just no longer exists, so don't worry! You can re-sample as many times as you want (re-run code).
- ✓ Please check the comments in the code for all details.

1. Code implementation and inference [3pt]

Clear the 'pass' of the written part like the picture attached below and **complete the code**. You should include a **loss graph** and **validation results** (3 images or so) of each file (RNN_Captiong.ipynb, LSTM_Captioning.ipynb, Transformer_Captioning.ipynb) in your report.



①. RNN_Captioning.ipynb

- To run this file, you need to complete the code in the **CV/rnn_layers.py**, **CV/classifiers/run.py** file. (Modifying the part related to RNN)

②. LSTM_Captioning.ipynb

- To run this file, you need to complete the code in the **CV/rnn_layers.py**, **CV/classifiers/run.py** file. (Modifying the part related to LSTM)

③. Transformer_Captioning.ipynb

- To run this file, you need to complete the code in the **CV/transformer_layer.py**, **CV/classifiers/transformer.py** file.

2. Implementation of evaluation matric (CIDEr score) [2pt]

Calculate and compare the CIDEr scores for RNN, LSTM, and Transformer results and attach them to the report. Refer to the link below to calculate the score of the training result and validation result by matching the input shape well.

(Reference link: CIDEr Code github)

3. Visualize the attention map. [1pt]

Visualize your attention scores using an attention map **for each word in the caption**. You can obtain the attention weights from **the MultiHeadAttention class** in CV/transformer_layers.py. Visualize these weights with matplotlib.pyplot, and include your results in the report.

4. Inference on NICE Challenge Dataset. [2pt]

Try to improve the performance using all methods that can improve the model performance (hyper-parameter modifications, ensemble techniques, etc.). However, it is prohibited to use additional data or use pre-trained models.

The descriptions for method to submit the results is on the website (click the hyperlink)

Those who significantly improve performance will receive better grades.

So, please do your best to enhance your performance.

** RECOMMENDED ** If you want to know if an idea to increase performance is appropriate, I strongly recommend contacting professor for feedback.