Assignment 7

Problem Statement

Report on COCO dataset and building basic image captioning model prototype.

Data:

Paper: https://arxiv.org/pdf/1405.0312.pdf

Data: https://www.kaggle.com/datasets/hariwh0/ms-coco-dataset

For images use folder with name train2014, and to extract the captions load captions/annotations/captions_train2014.json file

Task 1:

1. Prepare a summary report /literature review of the paper of minimum 2 pages.
2. Explain how data was prepared and what complexity/challenges were faced while preparing the data.
3. Recent developments in the COCO dataset and its model.

Note: Plagiarism will be checked. Individual task.

Task 2:

1. Create a dataframe with two columns, image filename and its associated captions from .json file.
2. Select any random 2000 as training and 500 as testing.
3. Use the Xception model to predict the feature matrix and dump it into the file.
4. Create the input sequences from caption using tokenization, and sequencing. (Refer Lab workflow)
5. Design the DNN-RNN model, with 2 LSTM layers, and 1 DNN layer for the Xception feature matrix.
   a. Add dropout and batch normalization layers.
6. Select random 12 images from the test set and display it with predicted captions.

Note: Task 2 can be submitted in a group of 5 students. Upload the notebook individually.