Setting the stage

Training a network for generating captions from images is quite a herculean task if one has to do everything from scratch.

We are blessed with a lot of predefined helper functions to easily implement a number of tasks. We are going to use them. Let's import them.

I've mentioned a comment as to why every import is crucial and what function it serves.

```
# To get the image and caption data from Google Drive.
from google.colab import drive

# For preprocessing captions and storing them in
# memory as dictionaries
from collections import defaultdict, Counter

# To navigate across folders and accessing data
import os

# To return a stack trace of errors if something goes wrong in GPU Computations
os.environ['CUDA_LAUNCH_BLOCKING'] = "1"

# For preprocessing captions
import string
import random
import re

# For matrix and math operations
import numpy as np
import math

# For formatting output while viewing training logs
import sys
import pandas as pd

# To compute the time required for training a
# batch, an epoch etc.
import time

# To build, train and save NN architectures for caption generation
import torch
import torchvision.transforms as tfms

# To define our own Dataset Class and dataloader instance
import torch.utils.data as data

# To read and resize images
from PIL import Image
```

Could not connect to the reCAPTCHA service. Please check your internet connection and reload to get a reCAPTCHA challenge.