IMAGE CAPTION GENERATOR

You saw an image and your brain can easily tell what the image is about, but can a computer tell what the image is representing? Computer vision researchers worked on this a lot and they considered it impossible until now! With the advancement in Deep learning techniques, availability of huge datasets and computer power, we can build models that can generate captions for an image.

This is what we are going to implement in this project where we will use combination of deep learning techniques - Convolutional Neural Networks and a type of Recurrent Neural Network (together).

What is Image Caption Generator?

Image caption generator is a task that involves computer vision and natural language processing concepts to recognize the context of an image and describe them in a natural language like English.

Image Caption Generator – About the Project

The objective of our project is to learn the concepts of a CNN and RNN-LSTM model and build a working model of Image caption generator by implementing CNN with LSTM.

In this project, we will be implementing the caption generator using CNN (Convolutional Neural Networks) and LSTM (Long short term memory).

NOTE: (Optional)

The image features can be extracted from **Xception** which is a Transfer learned CNN model trained on the imagenet dataset and then we can feed the features into the LSTM model which will be responsible for generating the image captions.

The Dataset for the Project

For the image caption generator, we will be using the Flickr_8K dataset. There are also other big datasets like Flickr_30K and MSCOCO dataset but it can take weeks just to train the network so we will be using a small Flickr8k dataset. The advantage of a huge dataset is that we can build better models.

The Flickr_8k_text folder contains file Flickr8k.token which is the main file of our dataset that contains image name and their respective captions separated by newline("\n").

Pre-requisites

This project requires good knowledge of Deep learning, Python, working on Jupyter notebooks, Keras library, Numpy, and ***Natural language processing***.
Make sure you have installed all the following necessary libraries:

- pip install tensorflow
- keras
- pillow
- numpy
- tqdm
- jupyterlab

What is CNN?

Convolutional Neural networks are specialized deep neural networks which can process the data that has input shape like a 2D matrix. Images are easily represented as a 2D matrix and CNN is very useful in working with images.

CNN is basically used for image classifications and identifying if an image is a bird, a plane or Superman, etc.

It scans images from left to right and top to bottom to pull out important features from the image and combines the feature to classify images. It can handle the images that have been translated, rotated, scaled and changes in perspective.

What is LSTM?

LSTM stands for Long short term memory, they are a type of RNN (recurrent neural network) which is well suited for sequence prediction problems. Based on the previous text, we can predict what the next word will be. It has proven itself effective from the traditional RNN by overcoming the limitations of RNN which had short term memory. LSTM can carry out relevant information throughout the processing of inputs and with a forget gate, it discards non-relevant information.

Image Caption Generator Model

So, to make our image caption generator model, we will be merging these architectures. It is also called a CNN-RNN model.

- CNN is used for extracting features from the image. You can use the pre-trained model Xception or You can create your own custom model.
- LSTM will use the information from CNN to help generate a description of the image.

Downloaded from dataset:

- **Flicker8k_Dataset –** Dataset folder which contains 8091 images.
- **Flickr_8k_text –** Dataset folder which contains text files and captions of images.

ALL THE BEST