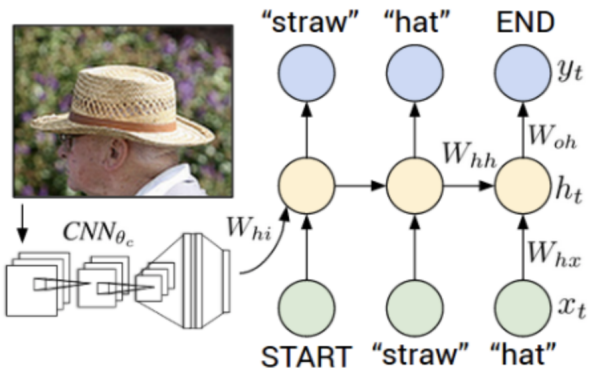
Prepare python notebook (recommended- use Google Colab) to build, train and evaluate model (tensorflow or tensorflow.keras library recommended). Read the instructions carefully.

Question: Image Captioning : Image Captioning is the process of generating textual description of an image. It uses both Natural Language Processing and Computer Vision to generate the captions. The dataset will be in the form [image → captions]. The dataset consists of input images and their corresponding output captions.



Encoder
The Convolutional Neural Network(CNN) can be thought of as an encoder. The input image is given to CNN to extract the features. The last hidden state of the CNN is connected to the Decoder.

Decoder
The Decoder is a Recurrent Neural Network(RNN) which does language modelling up to the word level. The first time step receives the encoded output from the encoder and also the <START> vector.

(13 marks)

1. **Import Libraries/Dataset (0 mark)**

a. Import the required libraries

b. Check the GPU available (recommended- use free GPU provided by Google Colab).

2. **Data Visualization and augmentation (3 mark)**

1. Read the pickle file (https://drive.google.com/file/d/1QG2_tAFLroc4ATSOLRiEAwYTYTsL_xVc/view?usp=sharing) and convert the data into the correct format which could be used for ML model.
   Pickle file contains the image id and the text associated with the image.
   Eg: '319847657_2c40e14113.jpg#0\tA girl in a purple shirt hold a pillow .
   Each image can have multiple captions.
   319847657_2c40e14113.jpg -> image name
   #0 -> Caption ID
   \t -> separator between Image name and Image Caption
   A girl in a purple shirt hold a pillow . -> Image Caption
   Corresponding image wrt image name can be found in the image dataset folder.

Image dataset Folder : https://drive.google.com/file/d/1-mPKMpphaKqtT26ZzbR5hCHGedkNyAf1/view?usp=sharing

a. Plot at least two samples and their captions (use matplotlib/seaborn/any other library).

b. Bring the train and test data in the required format.

3. **Model Building (7 mark)**

a. Use Pretrained VGG-16 model trained on ImageNet dataset (available publicly on google) for image feature extraction.

b. Create 3 layered GRU layer model and other relevant layers for image caption generation.

c. Add L2 regularization to all the GRU layers.

d. Add one layer of dropout at the appropriate position and give reasons.

e. Choose the appropriate activation function for all the layers.

f. Print the model summary.

4. **Model Compilation (1 mark)**

a. Compile the model with the appropriate loss function.

b. Use an appropriate optimizer. Give reasons for the choice of learning rate and its value.

5. **Model Training (1 mark)**

a. Train the model for an appropriate number of epochs. Print the train and validation loss for each epoch. Use the appropriate batch size.

b. Plot the loss and accuracy history graphs for both train and validation set. Print the total time taken for training.

6. **Model Evaluation (1 mark)**

a. Take a random image from google and generate caption for that image.