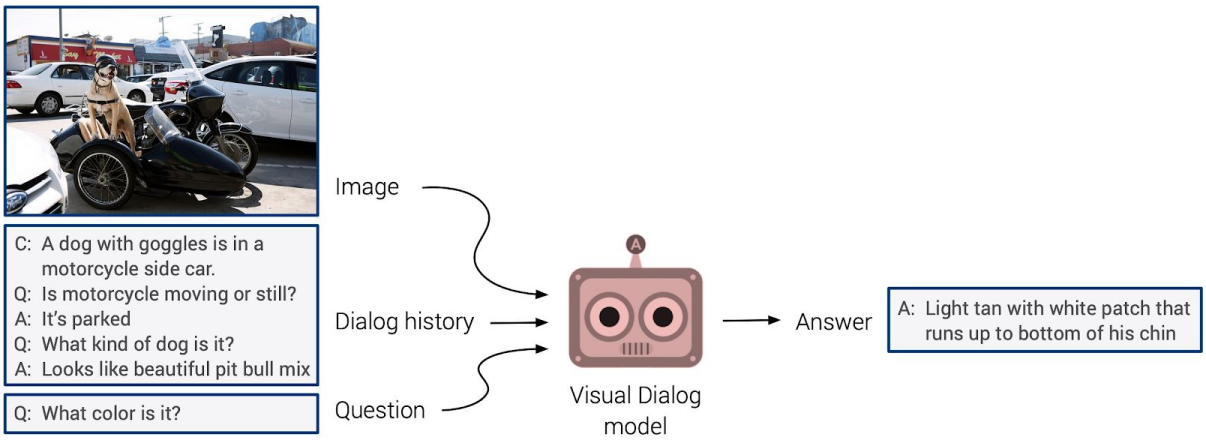
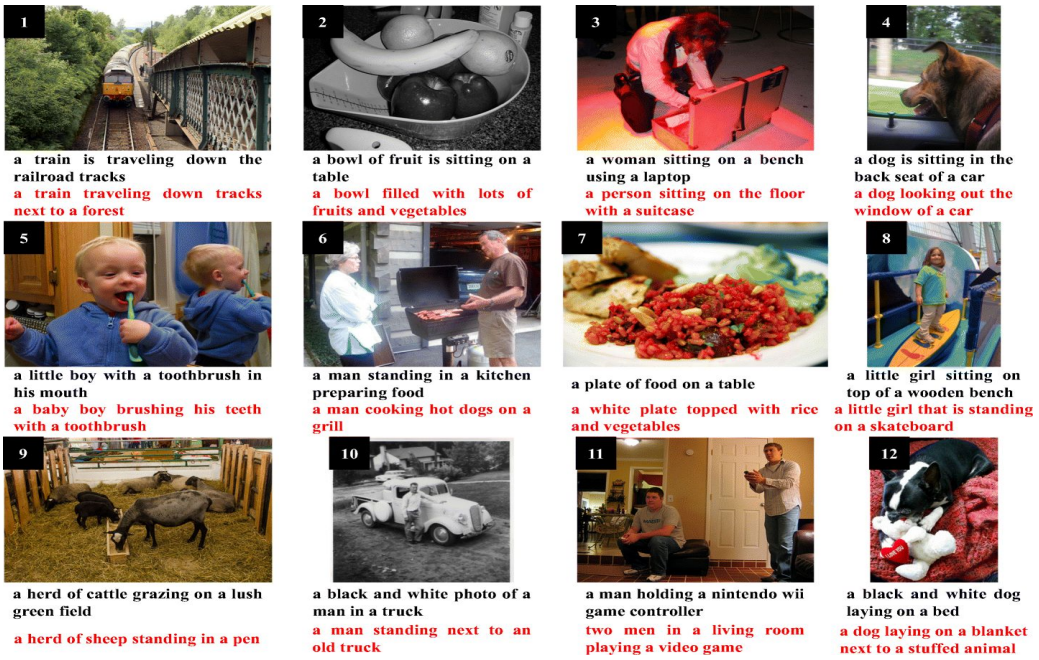
Project Name: “Answering from VisualDialog”

Purpost (Effect) of Project Goal

To predict an answer from input image and question using Long short term memory (LSTM) , Convolution neural network (CNN) and Attention weight.

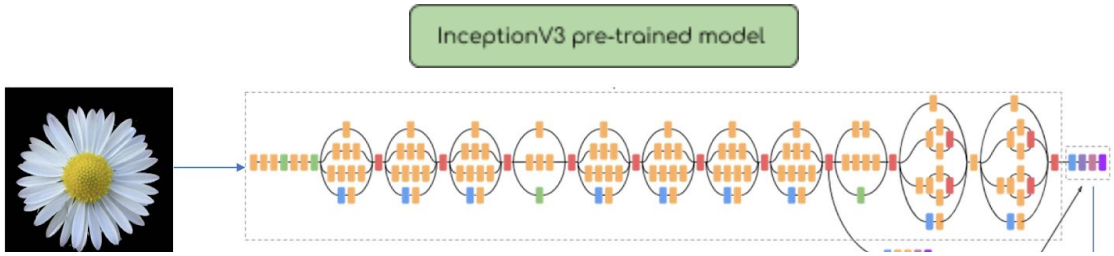


1. Download dataset of image question answer pair from https://visualdialog.org/data

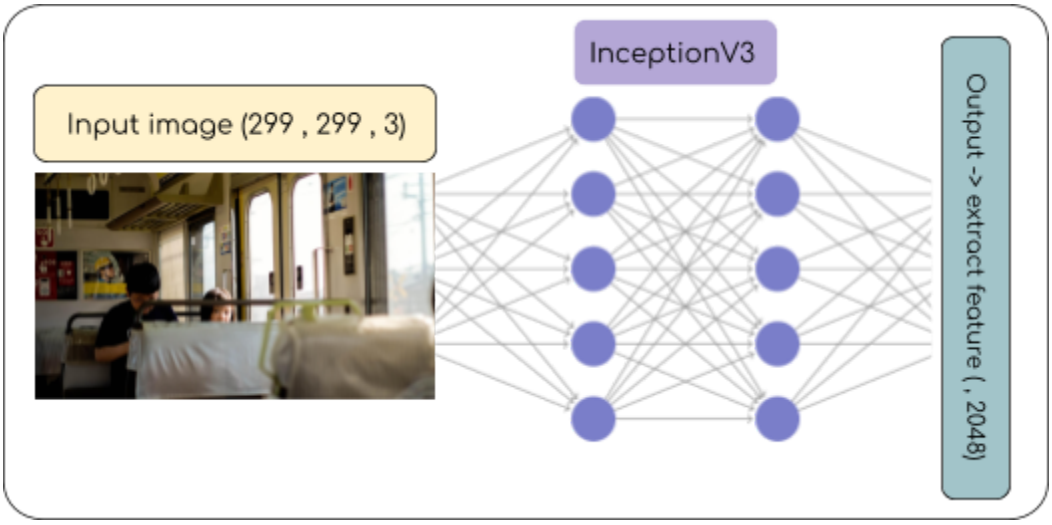


2. Extract feature of image using a transfer learning from pre-train model InceptionV3

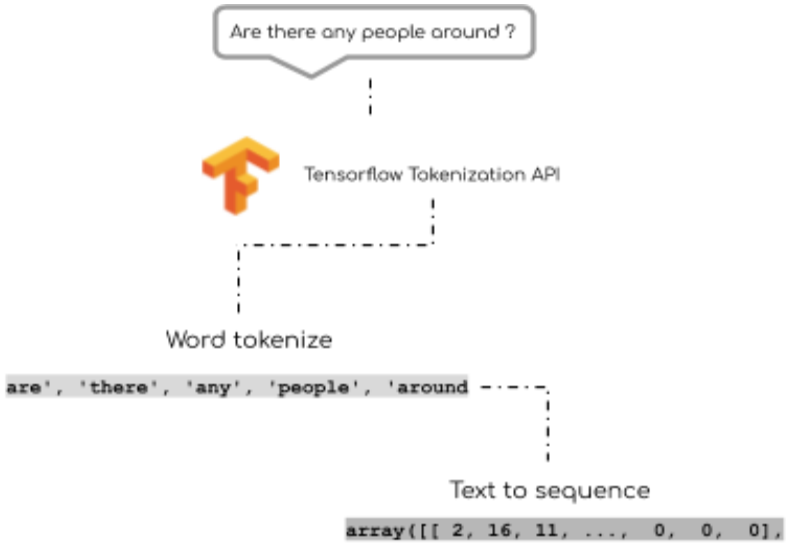
For Input be a first layer of Inception and output be a last layer of Inception.

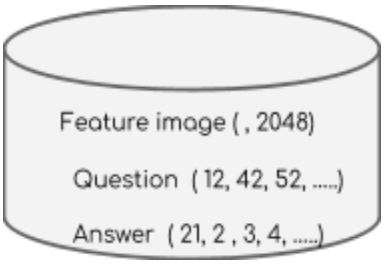


We will get the extract feature model for extract image to feature format



Preparing question and answer sentence using tensorflow Tokenization API , We will get an array of integer each number represent each word.

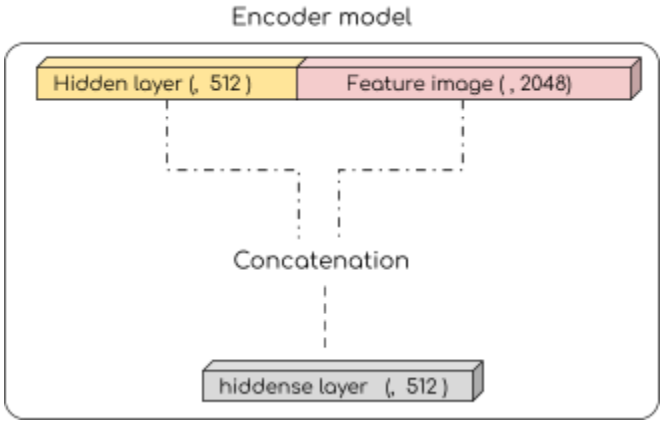


Create a Dataset which consist of 

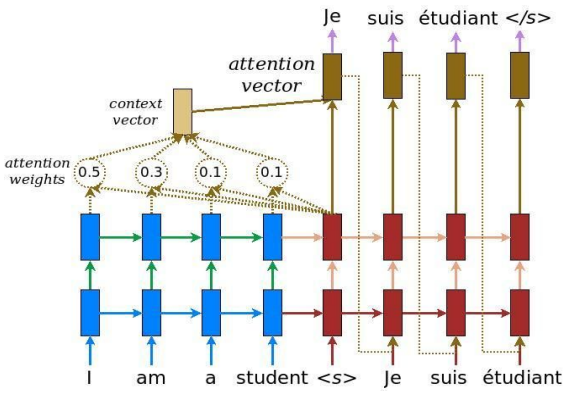
.

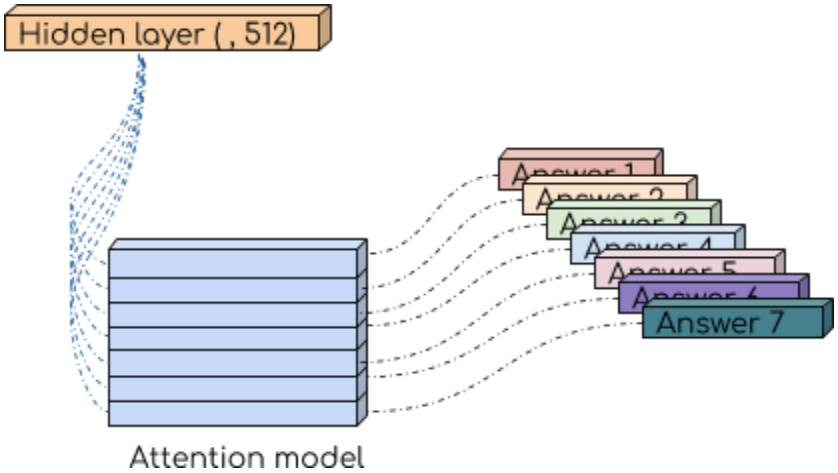
3. Create Encoder model , Attention weight model and Decoder model

For procedure of **Encoder model** is concatenation LSTM of question sentence and feature image to dense layer 512

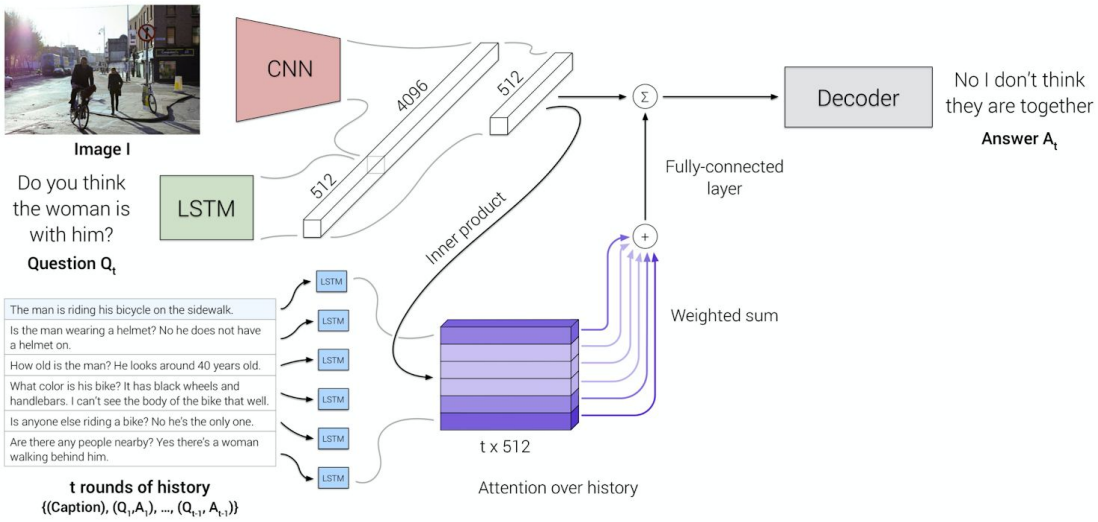


Attention model is that each input words is assigned a weight by the attention mechanism which is then used by the decoder to predict the next word in the sentence. The below picture and formulas are an example of attention mechanism from Luong's pape





Decoder model decode each answer word as target time by time



4. Training input is (feature image), (question word) target is (answer word) for 10 epoch and observed loss that should be decrease every epoch.

```
Epoch 9 Batch 0 Loss 0.0274
Epoch 9 Batch 100 Loss 0.0405
Epoch 9 Batch 200 Loss 0.0468
Epoch 9 Batch 300 Loss 0.0436
Epoch 9 Batch 400 Loss 0.0444
Epoch 9 Batch 500 Loss 0.0421
Epoch 9 Batch 600 Loss 0.0411
Epoch 9 Batch 700 Loss 0.0531
Epoch 9 Batch 800 Loss 0.0254
Epoch 9 Loss 0.0411
Time taken for 1 epoch 246.13827848434448 sec

Epoch 10 Batch 0 Loss 0.0380
Epoch 10 Batch 100 Loss 0.0260
Epoch 10 Batch 200 Loss 0.0292
Epoch 10 Batch 300 Loss 0.0558
Epoch 10 Batch 400 Loss 0.0543
Epoch 10 Batch 500 Loss 0.0406
Epoch 10 Batch 600 Loss 0.0477
Epoch 10 Batch 700 Loss 0.0458
Epoch 10 Batch 800 Loss 0.0509
Epoch 10 Loss 0.0395
Time taken for 1 epoch 247.42531752586365 sec
```

5. Testing model

```
Input: Question ['<start>', 'are', 'the', 'kites', 'pretty', '<end>']
Predicted Answer: i think so <end>
```


```
Input: Question ['<start>', 'is', 'it', 'night', 'time', '<end>']
Predicted Answer: it is daytime <end>
```