WORKSHOP ON COMPUTER VISION AND IMAGE PROCESSING
14 – 24 DECEMBER 2020

IMAGE CAPTIONING USING DEEP NEURAL NETWORK

By

Dr. M. MADHIARASAN

Post Doctoral Fellow,

Department of Computer Science and Engineering,

Indian Institute of Technology, Roorkee

Agenda

- What is Image Captioning?
- Why do we need Image Captioning and its Applications?
- Why do we need Deep Learning ?
- What is Deep Learning?
- What is Convolution Neural Networks?
- What is Recurrent Neural Network (LSTMs)?
- Design of Image Captioning Model using Deep Neural Network (CNN with LSTMs)
- Experiential Learning using Google Colab

What is Image Captioning?

Image Captioning is the process of automatically generating the context of the considered image with respect to the objects, the action happening in the image.

Considered Input Image :

Generated Caption of the Considered Image:

#1.A blonde horse and a blonde girl in a black sweatshirt are staring at a fire in a barrel

#2. A man , and girl and two horses are near a contained fire

Why do we need Image Captioning and its Applications?

Need : Aid the humanities and society

Applications:

❖**Image Searching Tool**

❖**Guidance Device**

❖**Self Driving Car**

❖**Traffic Signal**

❖**Surveillance and Security**

Why do we need Deep Learning ?

Need:

- **To perform feature extraction**
- **To perform complex operation**
- **To handle larger data**
- **To improve the performance with huge data set**

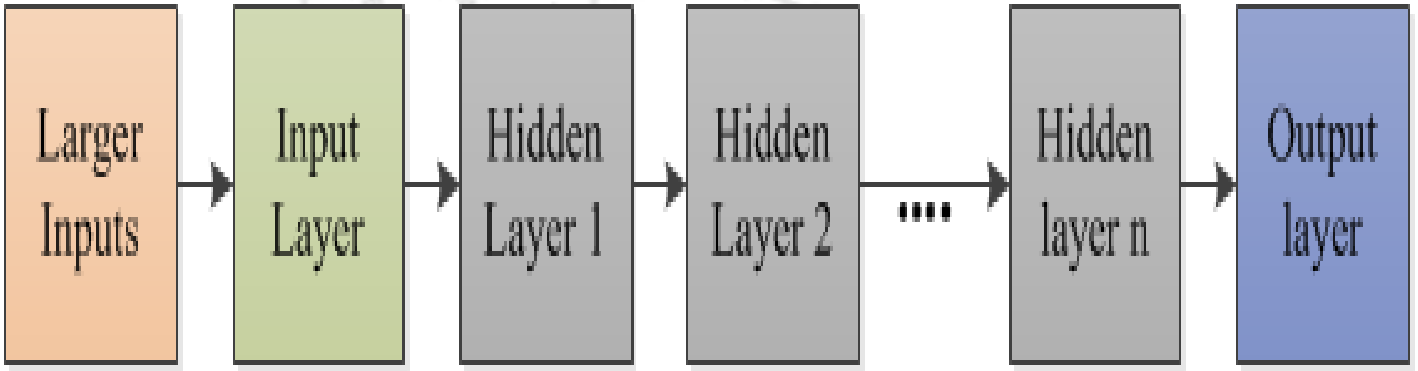
Applications:

- **Face Recognition**
- **Natural Language Processing**
- **Medical Diagnosis**
- **Digital Assistance**
- **Game Playing**
- **Speech Recognition**
- **Image Classification**
- **Hand Written Transcripts**
- **Self Driving**
- **Machine Translation**
- **Social Recommendation**
- **Surveillance and Security**

What is Deep Learning?

❖**Deep learning is a subset of Machine Learning (subset of AI) try to replicate the human brain structure using building of learning algorithms.**

❖**Extract the pattern from data using multiple hidden layer neural networks.**

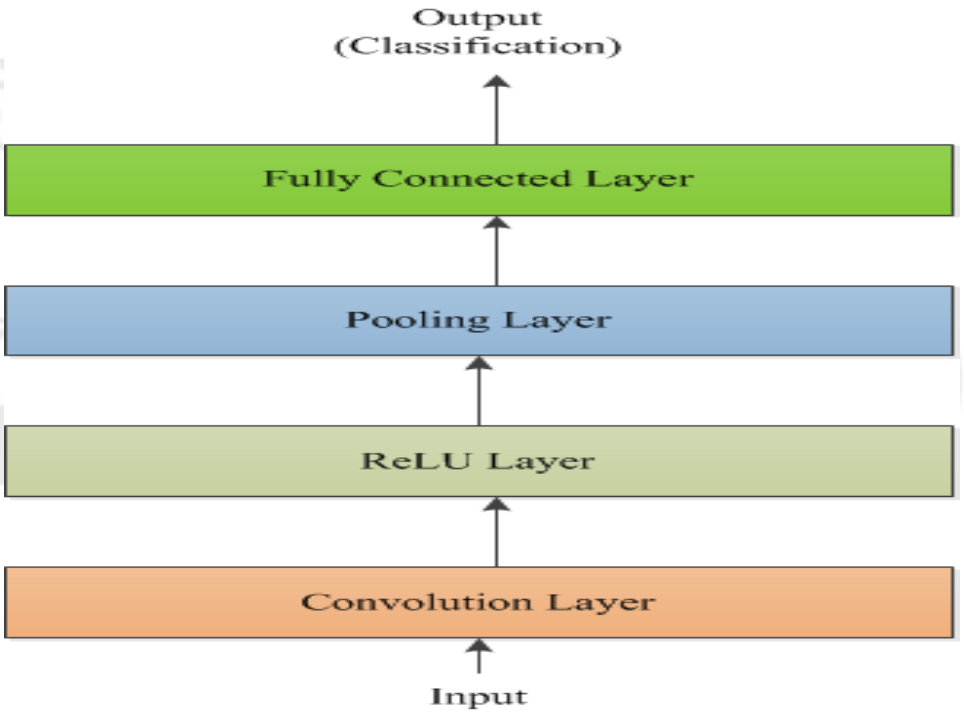


What is Convolution Neural Networks?

Why do we need CNNs?
ANNs requires more computation, memory and convergence takes more time

CNN is developed based on the inspiration of visual cortex, the key point of CNNs is a local understanding of an image is good enough.

General structure of CNNs

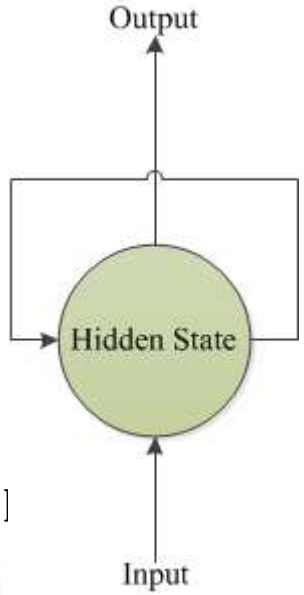


What is Recurrent Neural Network (LSTMs)?

Why do we need RNN?

- **In feed forward neural network output at 't' stamp is independent of the output at 't-1', so we can not predict the next words.**
- **To handle long term dependence.**

Recurrent Neural Network is developed for the purpose of capturing the information from the time series or sequence data.



Recurrent use BTT (Backpropagation Through Time) for training

Limitations of RNN:

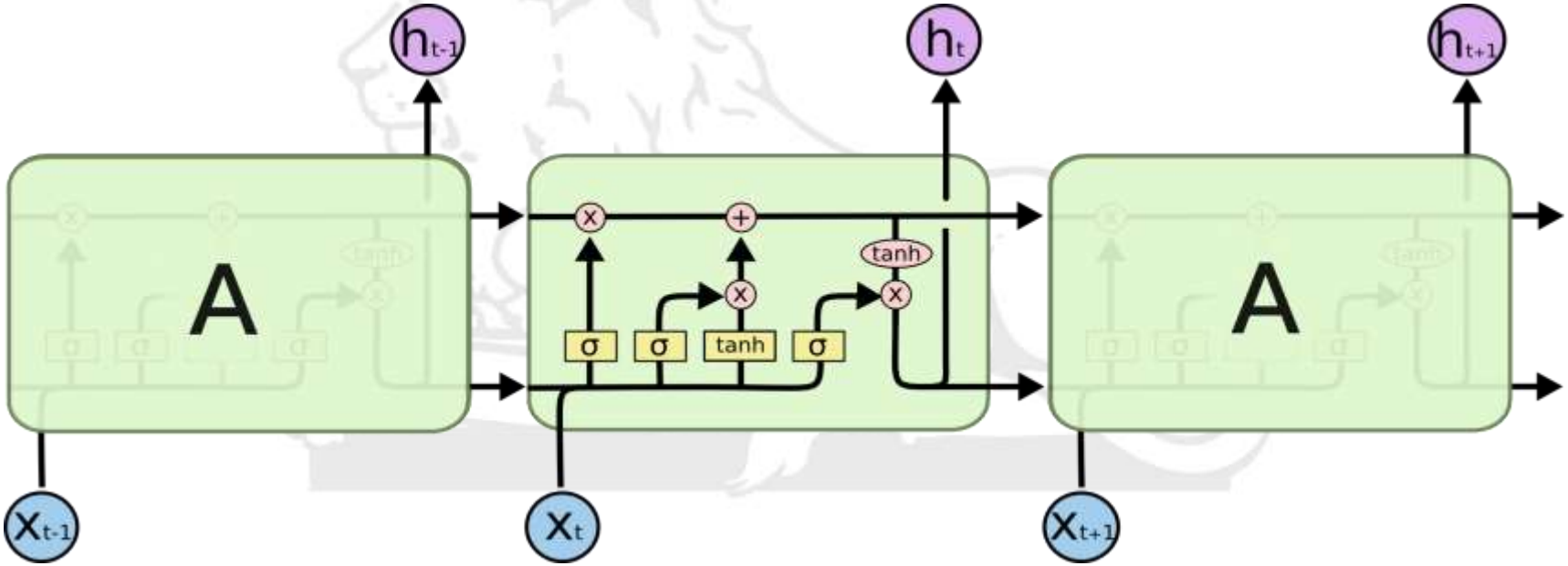
- **Exploding Gradients**
- **Vanishing Gradients**

LSTMs – Long Short Term Memory Networks

LSTMs is a special kind of RNN, it hold the past information for a longer period of time by means of memory cell.

Overcome the limitations of RNNs.

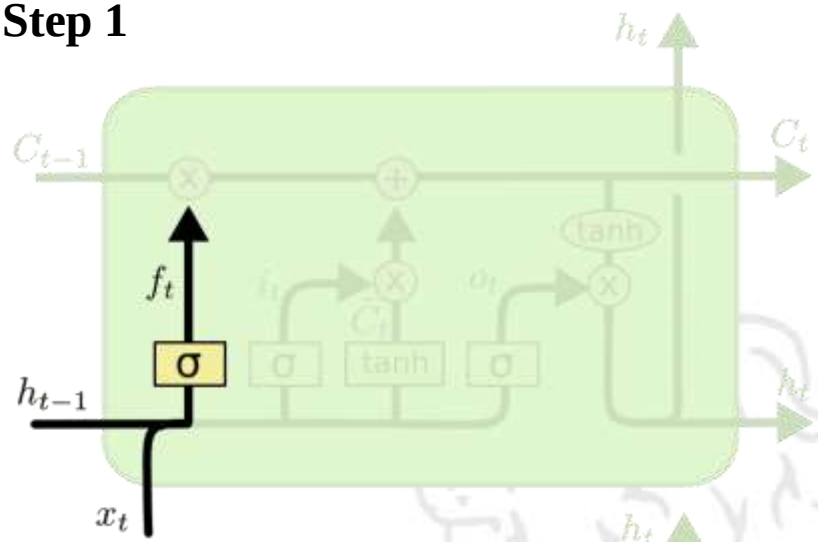
Posses learning long term dependence.



Source: http://colah.github.io/posts/2015-08-Understanding-LSTMs/

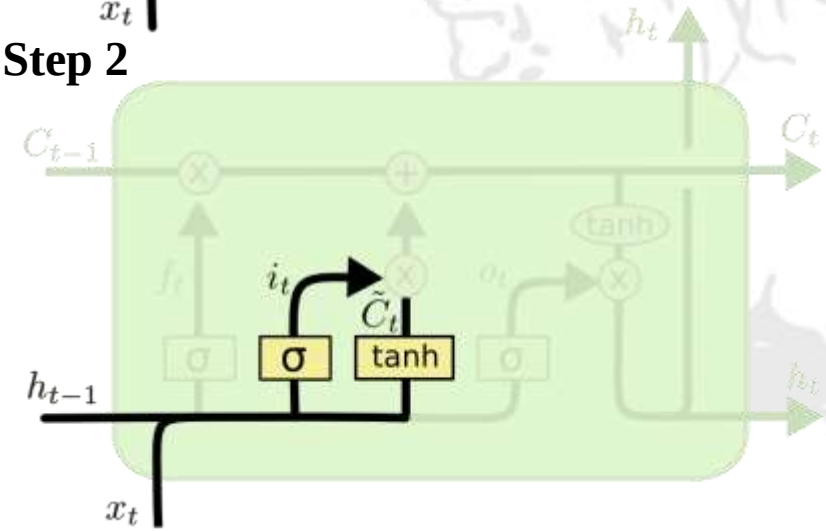
LSTMs – Long Short Term Memory Networks Continued

Step 1



$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \; + \; b_f\right)$$

Step 2



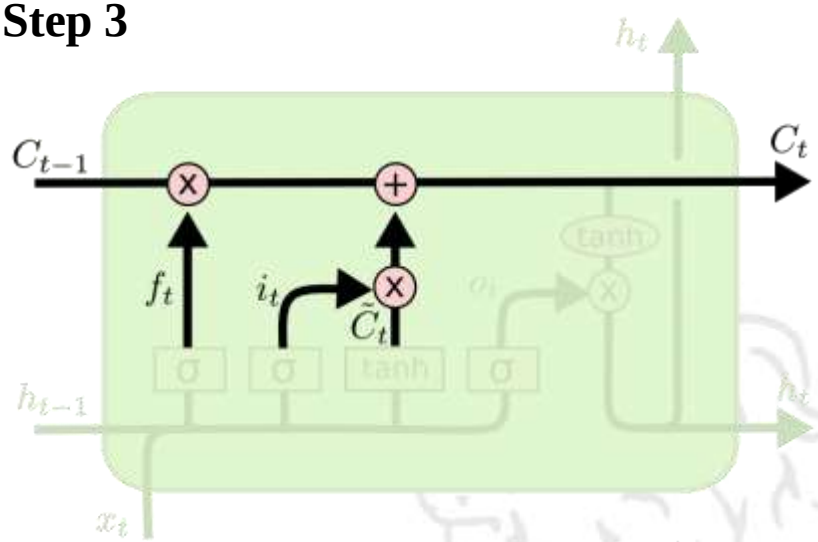
$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \; + \; b_i\right)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \; + \; b_C)$$

Source: http://colah.github.io/posts/2015-08-Understanding-LSTMs/

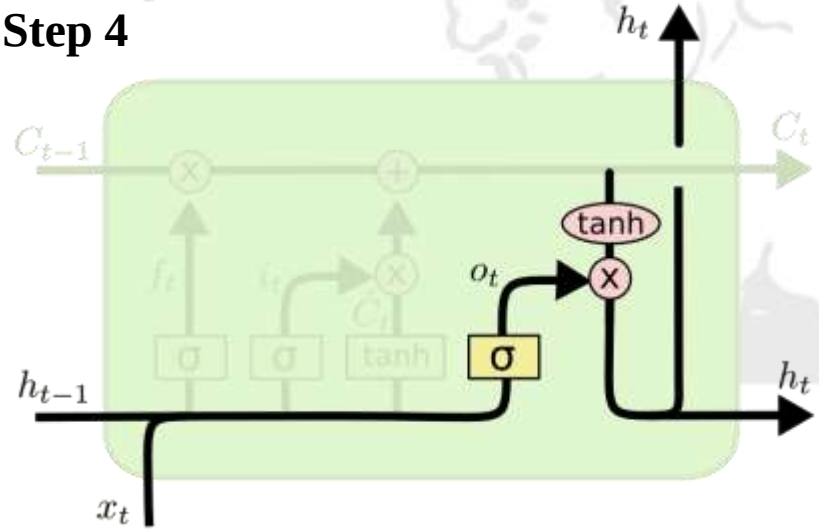
LSTMs – Long Short Term Memory Networks Continued

Step 3



$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

Step 4



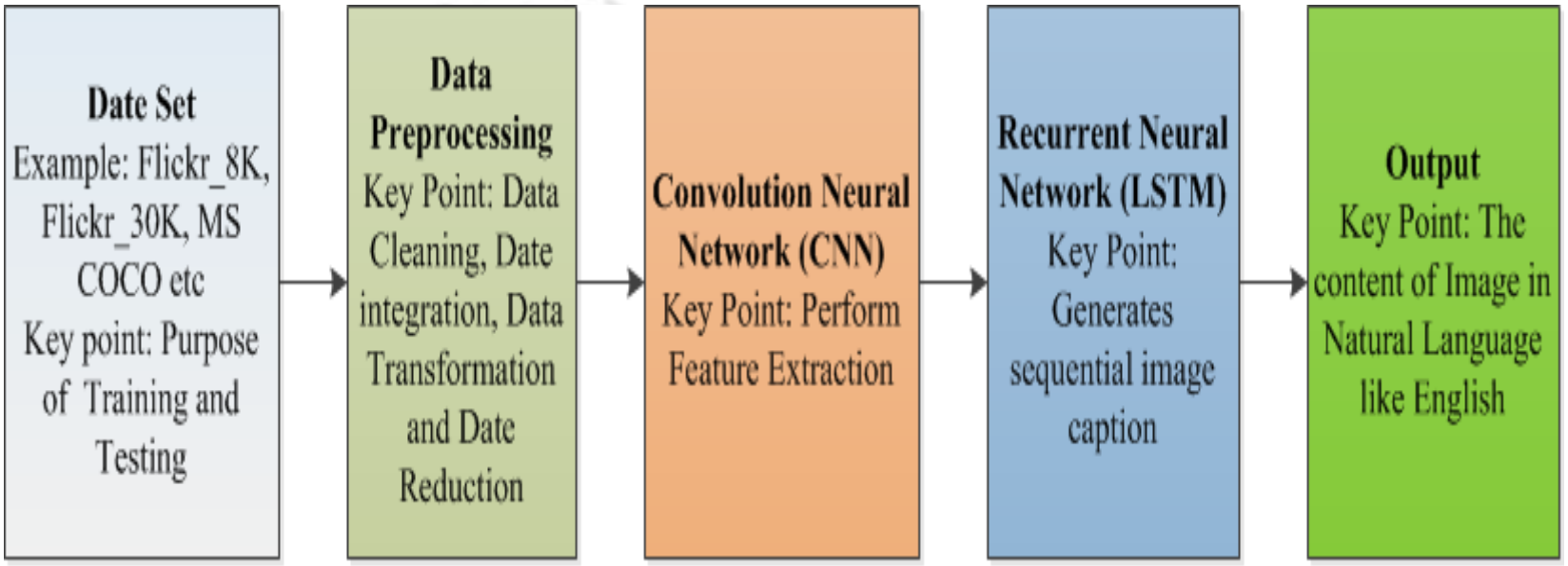
$$o_t = \sigma\left(W_o\left[h_{t-1}, x_t\right] + b_o\right)$$

$$h_t = o_t * \tanh\left(C_t\right)$$

Source: http://colah.github.io/posts/2015-08-Understanding-LSTMs/

Design of Image Captioning Model using Deep Neural Network (CNN with LSTMs)

The image captioning model designed based on the Convolution Neural Networks and Long Short Term Memory Networks



Block Diagram of Design of Image Captioning Model

Experiential Learning using Google Colab

The implementation of image captioning using deep neural network consist of following steps:

Step 1: Import all the necessary packages

Step 2: Load the data and perform the date preprocessing

Step 3: Extracting the feature vector from images using transfer learning (Xception Model)

Step 4: Load the preprocessed data to model for the purpose of training

Step5: Tokenizing the vocabulary

Step 6: Create data generator

Step 7: Define the structure of model (CNN –LSTMs)

Step 8: Perform training process

Step 9: Perform testing process

Step 10: Output

Implementation in Google Colab Platform

Image Captioning Using Deep Neural Network.ipynb

File Edit View Insert Runtime Tools Help Last saved at 6:47 AM

Comment Share

Files

drive

MyDrive

sample_data

Disk 75.83 GB available

+ Code + Text

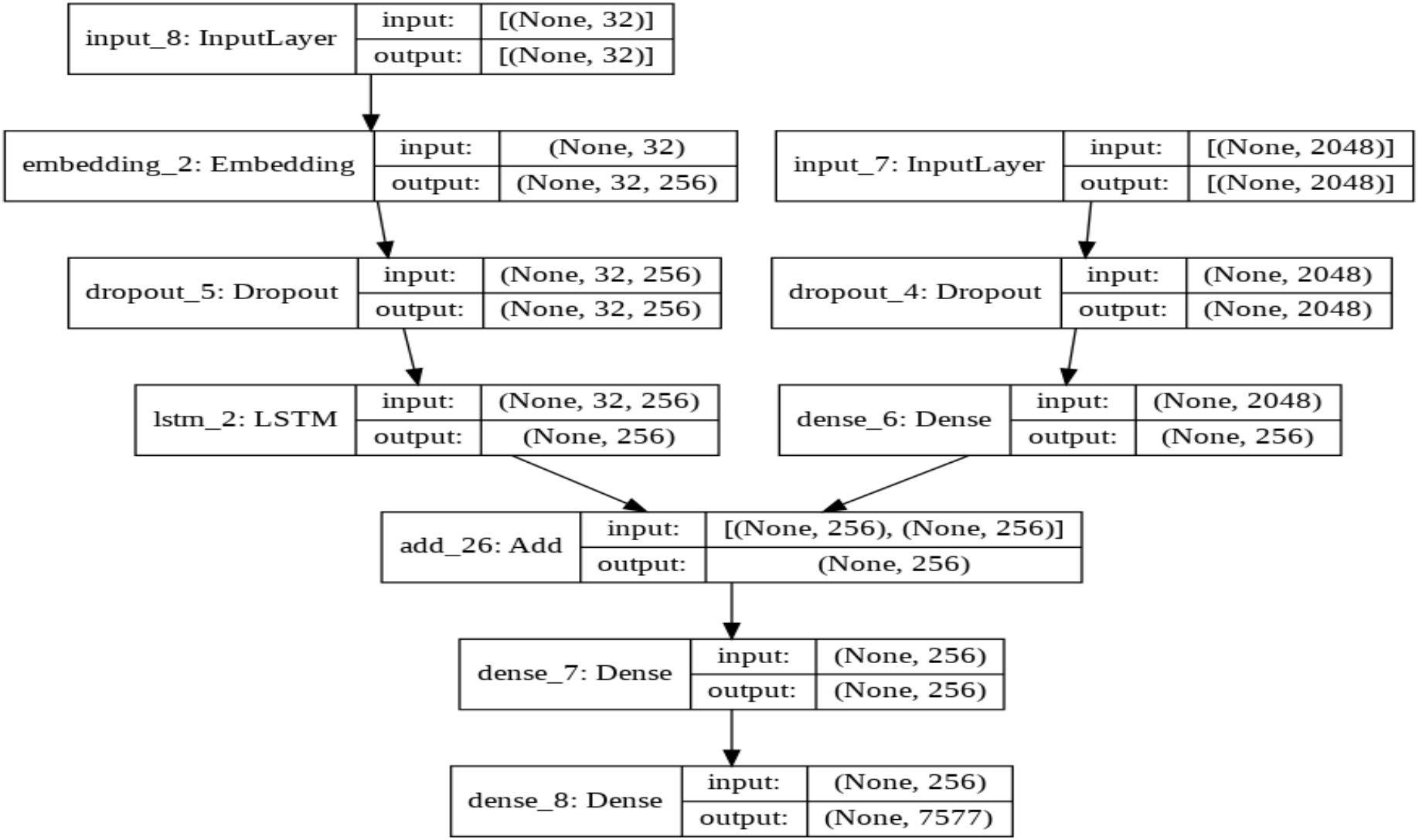
RAM Disk Editing

```
Workshop on Computer Vision and Image Processing
14-24 December 2020, IIT Roorkee

Image Captioning Using Deep Neural Network Tutorial Session
# \\ Dr. M. Madhiarasan, Post Doctoral Fellow \\
# \\ Mentor: Prof. Partha Pratim Roy \\
# \\ Computer Science and Engineering, IIT Roorkee \\
```

```
[ ] Note: Please Download the following files and upload in your Drive
    https://drive.google.com/drive/folders/1lJU4b3Uw-koRlEi97gfpVUbTJlA9Vo_i?usp=sharing
    https://drive.google.com/drive/folders/1_S1YGFhPNdDPRfvIYZXfVj1Y-7bzj7BN?usp=sharing
    https://drive.google.com/drive/folders/13zGHaBV7edOhESB3F4P4unokZwruB8bz?usp=sharing
```

```
[ ] import string
    import numpy as np
    from PIL import Image
    import os
    from pickle import dump, load
    import numpy as np

    from keras.applications.xception import Xception, preprocess_input
    from keras.preprocessing.image import load_img, img_to_array
    from keras.preprocessing.text import Tokenizer
    from keras.preprocessing.sequence import pad_sequences
    from keras.utils import to_categorical
    from keras.layers.merge import add
```

Designed Image Captioning Model



Designed Model Output

Image Captioning Using Deep Neural Network.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

+ Code + Text

RAM Disk Editing

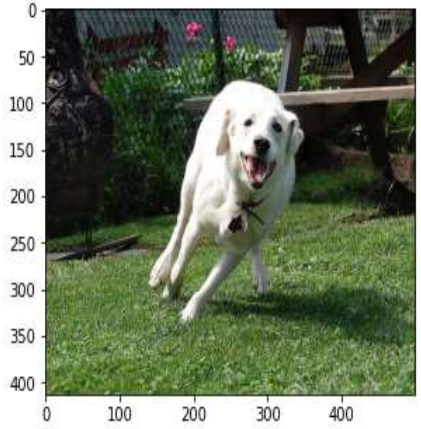
Files

- ..
- drive
  - MyDrive
    - Classroom
    - Colab Notebooks
      - Flickr8k_Dataset
      - Flickr8k_text
      - Test
        - 111537222_07e56...
        - 3457572788_e1fe4...
        - 3623302162_099f9...
        - 475778645_65b73...
        - COCO_test2015_00...
      - models
      - Copy of Image Recognit...
      - Image Captioning Using...
      - Image Recognition usin...
      - RNN.ipynb
      - Untitled0.ipynb
      - Untitled1.ipynb

Disk 75.62 GB available

```
img = image.open(img_path)

description = generate_desc(model, tokenizer, photo, max_length)
print("\n\n")
print(description)
plt.imshow(img)
```

```
Downloading data from https://storage.googleapis.com/tensorflow/keras-applications/xception/xception_weights_tf_dim_orderir
83689472/83683744 [==============================] - 1s 0us/step


start dog is running through the grass end
<matplotlib.image.AxesImage at 0x7f7c91da8f98>
```



Recent Trends in Image Captioning:

- Interactions Guided Generative Adversarial Network (IGGAN) for unsupervised image captioning
(Source:https://www.sciencedirect.com/science/article/pii/S0925231220312790)

- Variational Autoencoder and Reinforcement Learning for Remote sensing image captioning
(Source: https://www.sciencedirect.com/science/article/pii/S0950705120302586)

- DenseNet network and adaptive attention for Image captioning
(Source: https://www.sciencedirect.com/science/article/pii/S092359652030059X)

- Evolutionary recurrent neural network for image captioning
(Source: https://www.sciencedirect.com/science/article/pii/S0925231220304744)

- Multi-Level Policy and Reward-Based Deep Reinforcement Learning Framework for Image Captioning
(Source: https://ieeexplore.ieee.org/document/8844130)

Research Scope

- **Development of novel and automatic image captioning model**
- **Improve the neural network training**
- **Generates multiple captions accurately**
- **Working with open domain data set**
- **Design unsupervised image captioning model**

Where Do We Begin?

Finished!!! You did it!!!

Dr. M. Madhiarasan

Email ID: mmadhiarasan.cse@sric.iitr.ac.in