Lab3
Yue Zeng

I. Image Caption Generator by using show and tell

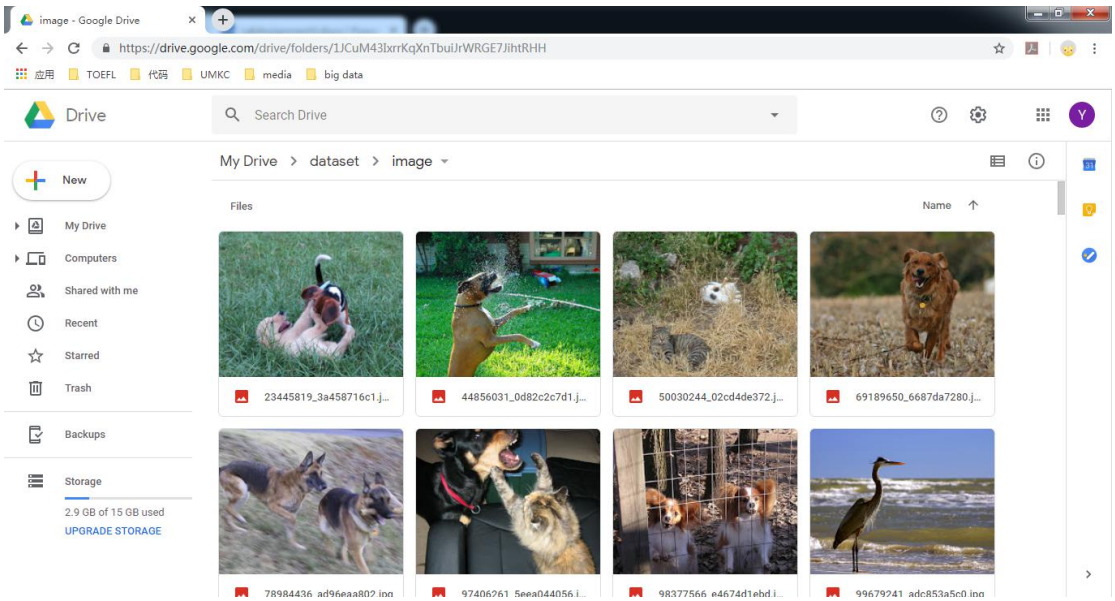
The show and tell model can recognize the object in the image, and show the relationship between them, then describe it using nature language. It is an encoder-decoder NN model. First it encodes the image to a representation, then it decodes the representation to a caption.("combine deep convolutional nets for image classification with recurrent networks for sequence modeling , to create a single network that generates descriptions of images"[1])

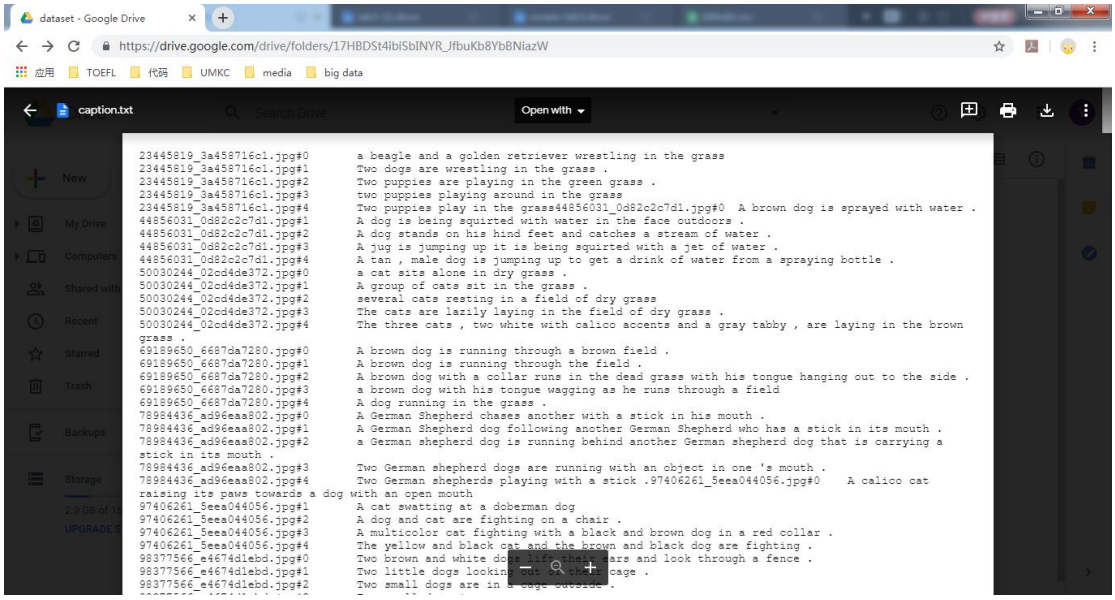
In the encoding step, it uses CNN, CNN can embed the image to fixed-length vector, this vector will became the input of the decoding step.

In the decoding step, it uses RNN with LSTM to generate the representation to natural language caption.
Machine translation: input a sentence, make the probability that the translation is correct maximal.
So we can use the same approach, input a image, use CNN to generate the object, and use RNN to "translate" it into description by maximizing the probability of the correction of description.

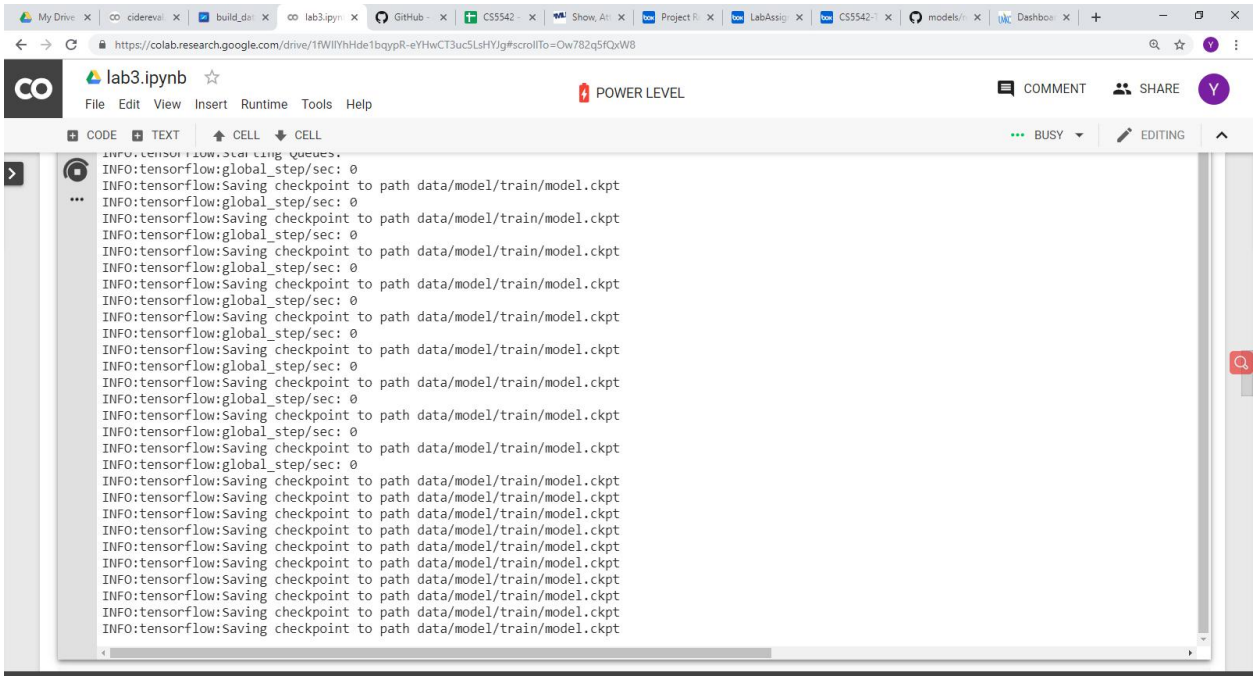
Since the theme of project is animals, we don't need to train the whole dataset for the model, so we select several images and the corresponding captions that match the theme to train the model.

1. Prepare the dataset for the model, including image and corresponding captions.





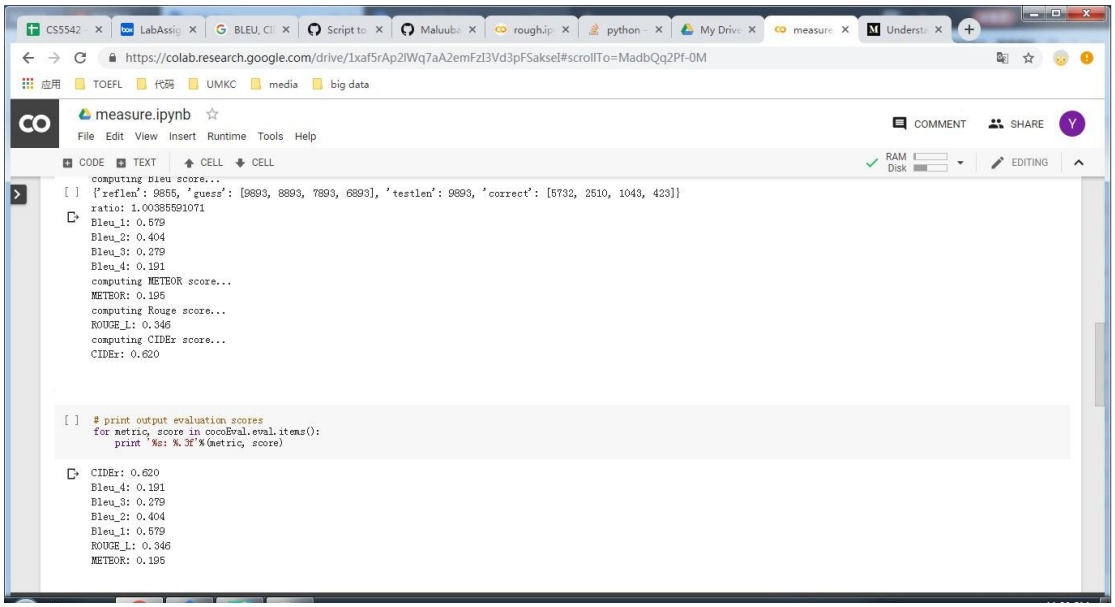
2. Train the model with the pre-trained checkpoint.



3. Generate caption for the test data

A blond dog runs down a flight of stairs to the backyard .
A dog jumps off the stairs .
A tan dog runs down a wooden staircase to the green grass .

4. accuracy in BLEU, CIDER, METEOR and ROGUE measures



BLEU-4: 0.191
CIDER: 0.620
METEOR: 0.196
ROGUE: 0.346

```
#The im2txt use the MSCOCO dataset, we use our own dataset.
#set dataset path
tf.flags.DEFINE_string("image_dir", "data/image/",
                       "Image directory.")
tf.flags.DEFINE_string("captions_file", "data/caption.txt",
                       "Captions text file.")
```

II. Unsupervised learning by using clustering

```
//Use the kmeans library of spark
import org.apache.spark.mllib.clustering.KMeans
val kMeansModel=KMeans.train(tf,10,1000) //model
val WSSSE = kMeansModel.computeCost(tf)//Within Set Sum of Squared Errors
val clusters=kMeansModel.predict(tf) //use predict function of the model
```

Output

A1 | 0

	A	B
1	0	
2	0	learning
3	0	Wikipedia the free encyclopedia
4	0	to navigationJump to search
5	0	deep vers see Stud see Artificial neural network.
6	0	learning and
7	0	mining
8	0	Machine.svg
9	0	
10	0	learning
11	0	◆ regression)
12	0	
13	0	
14	0	reduction[show]
15	0	prediction[show]
16	0	detection[show]
17	0	neural networks[show]
18	0	learning[show]
19	0	
20	0	venues[show]
21	0	of artificial intelligence[show]
22	0	articles[show]
23	0	Machine learning portal
24	0	
25	0	learning as oppos semi-supervised or unsupervised.[1][2][3]
26	0	

Reference

[1] Oriol Vinyals, Alexander Toshev, Samy Bengio, and Dumitru Erhan, “Show and Tell: Lessons learned from the 2015 MSCOCO Image Captioning Challenge”, in IEEE TRANSACTION ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, 2016