Final_Project_Image_Captioning_pc3019

December 9, 2022

0.0.1 Image Captioning

UNI: pc3019 By: Preethi Chandirasekeran

Another objective of this project was to perform image captioning of the book cover images. The image captioning model was trained using the Flickr8K dataset. It is a dataset of 8,000 images where each image is mapped to 5 captions which describe the content of the image. The image embedding was implemented using Inception v3 which is a model pretrained on ImageNet. It accepts an input shape of 299x299x3 and creates an embedding vector of dimensions 256x1. The captions from the dataset are preprocessed before passing to the RNNs. Text preprocessing includes converting to lowercase, removing special characters, and more. "startseq" is added to the beginning of each caption. Similarly, "endseq" is added to the end of each caption. The word embedding technique used is GloVe. GloVe is an unsupervised learning algorithm to obtain the word vector representations. The GloVe embedding representation used in this project has 6 billion tokens and 200 features. To obtain the predicted image caption, the image is passed to the model with the input string "startseq". Then the model predicts the next word and this word is appended to the input string. This repeats until "endseq" is reached or the maximum sentence length is reached.

[81]:
```
from time import time
import os
import glob
from tqdm import tqdm
import numpy as np
import tensorflow.keras as keras
import matplotlib.pyplot as plt
from tensorflow.keras.preprocessing import image
import string
import random
from tensorflow.keras.utils import to_categorical
from tensorflow.keras.layers import add
from tensorflow.keras.models import Model
from pickle import dump,load
from tensorflow.keras.preprocessing.sequence import pad_sequences
from tensorflow.keras.applications.inception_v3 import preprocess_input
from tensorflow.keras.applications.inception_v3 import InceptionV3
from tensorflow.keras.layers import Input,Dense,LSTM,Dropout,Embedding
from tensorflow.keras.utils import plot_model
```

[82]:
```
os.listdir('../input/flickr8k')
```

[82]:
```
['captions.txt', 'Images']
```

[83]:
```
my_images = glob.glob('../input/flickr8k/Images/' +'*.jpg')
```

[84]:
```
# Defining utility function which is used to load captions.txt
def loadFileUtil(file_name):
    f = open(file_name,'r')
    content = f.read()
    f.close()
    return content

captions = loadFileUtil('../input/flickr8k/captions.txt')
captions = captions.split('\n')
```

[85]:
```
captions.pop(0)
```

[85]:
```
'image,caption'
```

[86]:
```
captions.pop(-1)
```

[86]:
```
''
```

[87]:
```
# Formatting the captions
my_images = []
for x in range(len(captions)):
    curr_img = captions[x].split(',')
    curr_img = curr_img[0]
    my_images.append(curr_img)

my_images= set(my_images)
```

[88]:
```
# Populating a dictionary where the image name is the key and captions are the␣
 ↪value
imgNameDict = {}
for curr_img in my_images:
    imgNameDict[curr_img] = []

for curr_image_caption in captions:
    curr_image_caption = curr_image_caption.split(',')
    curr_img,mycap = curr_image_caption[0],curr_image_caption[-1]
    imgNameDict[curr_img].append(mycap)
```

[89]:
```
# Creating a translation table for punctuation removal
mytable = str.maketrans('', '', string.punctuation)
```

[90]:
```
for curr_img,captionList in imgNameDict.items():
    for x,curr_cap in enumerate(captionList):
        curr_cap = curr_cap.split()
```

```
        curr_cap = [y.lower() for y in curr_cap]

        curr_cap = [y.translate(mytable) for y in curr_cap]

        curr_cap = [y for y in curr_cap if len(y) >1]

        curr_cap = [y for y in curr_cap if y.isalpha()]

        captionList[x] = ' '.join(curr_cap)
```

[91]:
```
# Defining function to save the caption data
def saveCaptionData(caps,file_name):
    lines = []
    for img,capList in caps.items():
        for curr_cap in capList:
            lines.append(img + ',' + curr_cap)
        content = '\n'.join(lines)
        f =  open(file_name,'w')
        f.write(content)
        f.close()
```

[92]:
```
# Saving the cleaned captions
saveCaptionData(imgNameDict,'./cleanedCaptionData.txt')
```

[93]:
```
# Defining function to load the caption data
def loadCaptionData(file_name,image_path):
    myfile = loadFileUtil(file_name)
    myImageDict = {}
    my_images = glob.glob(image_path +'*.jpg')
    for idx,curr_image in enumerate(my_images):
        curr_image = curr_image.split('/')[-1]
        my_images[idx] = curr_image
        myImageDict[curr_image] = []
    for line in myfile.split('\n'):
        tokens = line.split(',')
        my_image= tokens[0]
        curr_cap=tokens[1:]
        if my_image in my_images:
            curr_cap = 'startseq ' + ' '.join(curr_cap) + ' endseq'
            myImageDict[my_image].append(curr_cap)
    return myImageDict

# Loading the cleaned caption data
```

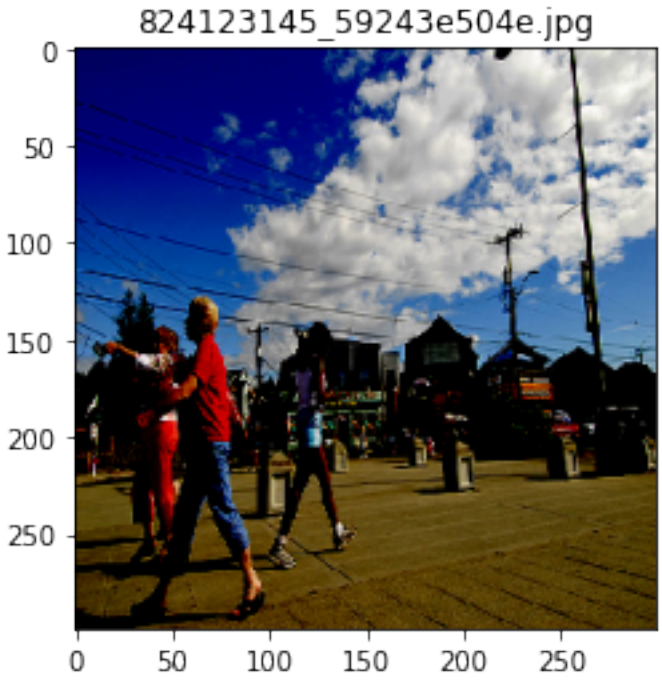
```
imgNameDict = loadCaptionData('./cleanedCaptionData.txt','../input/flickr8k/
 ↪Images/')
```

[94]:
```
# Defining utility function to preprocess the image
def preprocessingUtil(myPath):
    curr_img = image.load_img(myPath,target_size=  (299,299))
    temp= image.img_to_array(curr_img)
    temp = np.expand_dims(temp,axis =0)
    temp = preprocess_input(temp)
    return temp
```

[95]:
```
random_image = random.choice(list(my_images))
temp2  = preprocessingUtil('../input/flickr8k/Images/' + random_image)
for cap in imgNameDict[random_image]:
    print(cap)
plt.imshow(temp2[0])
plt.title(random_image)
```

```
startseq people walk outside on wooden walkway endseq
startseq three woman walk on the sidewalk endseq
startseq two elderly women are walking past younger woman on public path endseq
startseq walkers on concrete boardwalk under blue sky endseq
startseq women walking beneath blue sky and powerlines endseq
```

[95]:
```
Text(0.5, 1.0, '824123145_59243e504e.jpg')
```



[96]:
```
# Loading the InceptionV3 model
mymodel = InceptionV3(weights = 'imagenet')
```

[97]:
```
# Removing the last layer
featExtractor = Model(mymodel.input,mymodel.layers[-2].output)
featExtractor.summary()
```

```
Model: "model_2"
_________________________________________________________________________________
__________________
Layer (type)                    Output Shape         Param #     Connected to
=================================================================================
==================
input_4 (InputLayer)            [(None, 299, 299, 3) 0
_________________________________________________________________________________
__________________
conv2d_94 (Conv2D)              (None, 149, 149, 32) 864         input_4[0][0]
_________________________________________________________________________________
__________________
batch_normalization_94 (BatchNo (None, 149, 149, 32) 96          conv2d_94[0][0]
_________________________________________________________________________________
__________________
activation_94 (Activation)      (None, 149, 149, 32) 0
batch_normalization_94[0][0]
_________________________________________________________________________________
__________________
conv2d_95 (Conv2D)              (None, 147, 147, 32) 9216
activation_94[0][0]
_________________________________________________________________________________
__________________
batch_normalization_95 (BatchNo (None, 147, 147, 32) 96          conv2d_95[0][0]
_________________________________________________________________________________
__________________
activation_95 (Activation)      (None, 147, 147, 32) 0
batch_normalization_95[0][0]
_________________________________________________________________________________
__________________
conv2d_96 (Conv2D)              (None, 147, 147, 64) 18432
activation_95[0][0]
_________________________________________________________________________________
__________________
batch_normalization_96 (BatchNo (None, 147, 147, 64) 192         conv2d_96[0][0]
_________________________________________________________________________________
__________________
activation_96 (Activation)      (None, 147, 147, 64) 0
batch_normalization_96[0][0]
_________________________________________________________________________________
__________________
```

```
max_pooling2d_4 (MaxPooling2D)  (None, 73, 73, 64)   0
activation_96[0][0]
__________________________________________________________________________________
__________________
conv2d_97 (Conv2D)              (None, 73, 73, 80)   5120
max_pooling2d_4[0][0]
__________________________________________________________________________________
__________________
batch_normalization_97 (BatchNo (None, 73, 73, 80)   240         conv2d_97[0][0]
__________________________________________________________________________________
__________________
activation_97 (Activation)      (None, 73, 73, 80)   0
batch_normalization_97[0][0]
__________________________________________________________________________________
__________________
conv2d_98 (Conv2D)              (None, 71, 71, 192)  138240
activation_97[0][0]
__________________________________________________________________________________
__________________
batch_normalization_98 (BatchNo (None, 71, 71, 192)  576         conv2d_98[0][0]
__________________________________________________________________________________
__________________
activation_98 (Activation)      (None, 71, 71, 192)  0
batch_normalization_98[0][0]
__________________________________________________________________________________
__________________
max_pooling2d_5 (MaxPooling2D)  (None, 35, 35, 192)  0
activation_98[0][0]
__________________________________________________________________________________
__________________
conv2d_102 (Conv2D)             (None, 35, 35, 64)   12288
max_pooling2d_5[0][0]
__________________________________________________________________________________
__________________
batch_normalization_102 (BatchN (None, 35, 35, 64)   192
conv2d_102[0][0]
__________________________________________________________________________________
__________________
activation_102 (Activation)     (None, 35, 35, 64)   0
batch_normalization_102[0][0]
__________________________________________________________________________________
__________________
conv2d_100 (Conv2D)             (None, 35, 35, 48)   9216
max_pooling2d_5[0][0]
__________________________________________________________________________________
__________________
conv2d_103 (Conv2D)             (None, 35, 35, 96)   55296
activation_102[0][0]
```

```
____________________________________________________________________________
__________________
batch_normalization_100 (BatchN (None, 35, 35, 48)   144
conv2d_100[0][0]
____________________________________________________________________________
__________________
batch_normalization_103 (BatchN (None, 35, 35, 96)   288
conv2d_103[0][0]
____________________________________________________________________________
__________________
activation_100 (Activation)     (None, 35, 35, 48)   0
batch_normalization_100[0][0]
____________________________________________________________________________
__________________
activation_103 (Activation)     (None, 35, 35, 96)   0
batch_normalization_103[0][0]
____________________________________________________________________________
__________________
average_pooling2d_9 (AveragePoo (None, 35, 35, 192)  0
max_pooling2d_5[0][0]
____________________________________________________________________________
__________________
conv2d_99 (Conv2D)              (None, 35, 35, 64)   12288
max_pooling2d_5[0][0]
____________________________________________________________________________
__________________
conv2d_101 (Conv2D)             (None, 35, 35, 64)   76800
activation_100[0][0]
____________________________________________________________________________
__________________
conv2d_104 (Conv2D)             (None, 35, 35, 96)   82944
activation_103[0][0]
____________________________________________________________________________
__________________
conv2d_105 (Conv2D)             (None, 35, 35, 32)   6144
average_pooling2d_9[0][0]
____________________________________________________________________________
__________________
batch_normalization_99 (BatchNo (None, 35, 35, 64)   192         conv2d_99[0][0]
____________________________________________________________________________
__________________
batch_normalization_101 (BatchN (None, 35, 35, 64)   192
conv2d_101[0][0]
____________________________________________________________________________
__________________
batch_normalization_104 (BatchN (None, 35, 35, 96)   288
conv2d_104[0][0]
____________________________________________________________________________
```

```
------------------
batch_normalization_105 (BatchN (None, 35, 35, 32)   96
conv2d_105[0][0]
--------------------------------------------------------------------------------
------------------
activation_99 (Activation)      (None, 35, 35, 64)   0
batch_normalization_99[0][0]
--------------------------------------------------------------------------------
------------------
activation_101 (Activation)     (None, 35, 35, 64)   0
batch_normalization_101[0][0]
--------------------------------------------------------------------------------
------------------
activation_104 (Activation)     (None, 35, 35, 96)   0
batch_normalization_104[0][0]
--------------------------------------------------------------------------------
------------------
activation_105 (Activation)     (None, 35, 35, 32)   0
batch_normalization_105[0][0]
--------------------------------------------------------------------------------
------------------
mixed0 (Concatenate)            (None, 35, 35, 256)  0
activation_99[0][0]
activation_101[0][0]
activation_104[0][0]
activation_105[0][0]
--------------------------------------------------------------------------------
------------------
conv2d_109 (Conv2D)             (None, 35, 35, 64)   16384       mixed0[0][0]
--------------------------------------------------------------------------------
------------------
batch_normalization_109 (BatchN (None, 35, 35, 64)   192
conv2d_109[0][0]
--------------------------------------------------------------------------------
------------------
activation_109 (Activation)     (None, 35, 35, 64)   0
batch_normalization_109[0][0]
--------------------------------------------------------------------------------
------------------
conv2d_107 (Conv2D)             (None, 35, 35, 48)   12288       mixed0[0][0]
--------------------------------------------------------------------------------
------------------
conv2d_110 (Conv2D)             (None, 35, 35, 96)   55296
activation_109[0][0]
--------------------------------------------------------------------------------
------------------
batch_normalization_107 (BatchN (None, 35, 35, 48)   144
conv2d_107[0][0]
```

```
__________________________________________________________________________________
__________________
batch_normalization_110 (BatchN (None, 35, 35, 96)   288
conv2d_110[0][0]
__________________________________________________________________________________
__________________
activation_107 (Activation)     (None, 35, 35, 48)   0
batch_normalization_107[0][0]
__________________________________________________________________________________
__________________
activation_110 (Activation)     (None, 35, 35, 96)   0
batch_normalization_110[0][0]
__________________________________________________________________________________
__________________
average_pooling2d_10 (AveragePo (None, 35, 35, 256)  0           mixed0[0][0]
__________________________________________________________________________________
__________________
conv2d_106 (Conv2D)             (None, 35, 35, 64)   16384       mixed0[0][0]
__________________________________________________________________________________
__________________
conv2d_108 (Conv2D)             (None, 35, 35, 64)   76800
activation_107[0][0]
__________________________________________________________________________________
__________________
conv2d_111 (Conv2D)             (None, 35, 35, 96)   82944
activation_110[0][0]
__________________________________________________________________________________
__________________
conv2d_112 (Conv2D)             (None, 35, 35, 64)   16384
average_pooling2d_10[0][0]
__________________________________________________________________________________
__________________
batch_normalization_106 (BatchN (None, 35, 35, 64)   192
conv2d_106[0][0]
__________________________________________________________________________________
__________________
batch_normalization_108 (BatchN (None, 35, 35, 64)   192
conv2d_108[0][0]
__________________________________________________________________________________
__________________
batch_normalization_111 (BatchN (None, 35, 35, 96)   288
conv2d_111[0][0]
__________________________________________________________________________________
__________________
batch_normalization_112 (BatchN (None, 35, 35, 64)   192
conv2d_112[0][0]
__________________________________________________________________________________
__________________
```

```
activation_106 (Activation)     (None, 35, 35, 64)   0
batch_normalization_106[0][0]
__________________________________________________________________________________
__________________
activation_108 (Activation)     (None, 35, 35, 64)   0
batch_normalization_108[0][0]
__________________________________________________________________________________
__________________
activation_111 (Activation)     (None, 35, 35, 96)   0
batch_normalization_111[0][0]
__________________________________________________________________________________
__________________
activation_112 (Activation)     (None, 35, 35, 64)   0
batch_normalization_112[0][0]
__________________________________________________________________________________
__________________
mixed1 (Concatenate)            (None, 35, 35, 288)  0
activation_106[0][0]
activation_108[0][0]
activation_111[0][0]
activation_112[0][0]
__________________________________________________________________________________
__________________
conv2d_116 (Conv2D)             (None, 35, 35, 64)   18432       mixed1[0][0]
__________________________________________________________________________________
__________________
batch_normalization_116 (BatchN (None, 35, 35, 64)   192
conv2d_116[0][0]
__________________________________________________________________________________
__________________
activation_116 (Activation)     (None, 35, 35, 64)   0
batch_normalization_116[0][0]
__________________________________________________________________________________
__________________
conv2d_114 (Conv2D)             (None, 35, 35, 48)   13824       mixed1[0][0]
__________________________________________________________________________________
__________________
conv2d_117 (Conv2D)             (None, 35, 35, 96)   55296
activation_116[0][0]
__________________________________________________________________________________
__________________
batch_normalization_114 (BatchN (None, 35, 35, 48)   144
conv2d_114[0][0]
__________________________________________________________________________________
__________________
batch_normalization_117 (BatchN (None, 35, 35, 96)   288
conv2d_117[0][0]
__________________________________________________________________________________
```

```
__________________
activation_114 (Activation)     (None, 35, 35, 48)   0
batch_normalization_114[0][0]

__________________________________________________________________________________
__________________
activation_117 (Activation)     (None, 35, 35, 96)   0
batch_normalization_117[0][0]

__________________________________________________________________________________
__________________
average_pooling2d_11 (AveragePo (None, 35, 35, 288)  0           mixed1[0][0]

__________________________________________________________________________________
__________________
conv2d_113 (Conv2D)             (None, 35, 35, 64)   18432       mixed1[0][0]

__________________________________________________________________________________
__________________
conv2d_115 (Conv2D)             (None, 35, 35, 64)   76800
activation_114[0][0]

__________________________________________________________________________________
__________________
conv2d_118 (Conv2D)             (None, 35, 35, 96)   82944
activation_117[0][0]

__________________________________________________________________________________
__________________
conv2d_119 (Conv2D)             (None, 35, 35, 64)   18432
average_pooling2d_11[0][0]

__________________________________________________________________________________
__________________
batch_normalization_113 (BatchN (None, 35, 35, 64)   192
conv2d_113[0][0]

__________________________________________________________________________________
__________________
batch_normalization_115 (BatchN (None, 35, 35, 64)   192
conv2d_115[0][0]

__________________________________________________________________________________
__________________
batch_normalization_118 (BatchN (None, 35, 35, 96)   288
conv2d_118[0][0]

__________________________________________________________________________________
__________________
batch_normalization_119 (BatchN (None, 35, 35, 64)   192
conv2d_119[0][0]

__________________________________________________________________________________
__________________
activation_113 (Activation)     (None, 35, 35, 64)   0
batch_normalization_113[0][0]

__________________________________________________________________________________
__________________
activation_115 (Activation)     (None, 35, 35, 64)   0
```

```
batch_normalization_115[0][0]
__________________________________________________________________________________
__________________
activation_118 (Activation)     (None, 35, 35, 96)   0
batch_normalization_118[0][0]
__________________________________________________________________________________
__________________
activation_119 (Activation)     (None, 35, 35, 64)   0
batch_normalization_119[0][0]
__________________________________________________________________________________
__________________
mixed2 (Concatenate)            (None, 35, 35, 288)  0
activation_113[0][0]
activation_115[0][0]
activation_118[0][0]
activation_119[0][0]
__________________________________________________________________________________
__________________
conv2d_121 (Conv2D)             (None, 35, 35, 64)   18432       mixed2[0][0]
__________________________________________________________________________________
__________________
batch_normalization_121 (BatchN (None, 35, 35, 64)   192
conv2d_121[0][0]
__________________________________________________________________________________
__________________
activation_121 (Activation)     (None, 35, 35, 64)   0
batch_normalization_121[0][0]
__________________________________________________________________________________
__________________
conv2d_122 (Conv2D)             (None, 35, 35, 96)   55296
activation_121[0][0]
__________________________________________________________________________________
__________________
batch_normalization_122 (BatchN (None, 35, 35, 96)   288
conv2d_122[0][0]
__________________________________________________________________________________
__________________
activation_122 (Activation)     (None, 35, 35, 96)   0
batch_normalization_122[0][0]
__________________________________________________________________________________
__________________
conv2d_120 (Conv2D)             (None, 17, 17, 384)  995328      mixed2[0][0]
__________________________________________________________________________________
__________________
conv2d_123 (Conv2D)             (None, 17, 17, 96)   82944
activation_122[0][0]
__________________________________________________________________________________
__________________
```

```
batch_normalization_120 (BatchN (None, 17, 17, 384)  1152
conv2d_120[0][0]
__________________________________________________________________________________
__________________
batch_normalization_123 (BatchN (None, 17, 17, 96)   288
conv2d_123[0][0]
__________________________________________________________________________________
__________________
activation_120 (Activation)     (None, 17, 17, 384)  0
batch_normalization_120[0][0]
__________________________________________________________________________________
__________________
activation_123 (Activation)     (None, 17, 17, 96)   0
batch_normalization_123[0][0]
__________________________________________________________________________________
__________________
max_pooling2d_6 (MaxPooling2D)  (None, 17, 17, 288)  0           mixed2[0][0]
__________________________________________________________________________________
__________________
mixed3 (Concatenate)            (None, 17, 17, 768)  0
activation_120[0][0]
activation_123[0][0]
max_pooling2d_6[0][0]
__________________________________________________________________________________
__________________
conv2d_128 (Conv2D)             (None, 17, 17, 128)  98304       mixed3[0][0]
__________________________________________________________________________________
__________________
batch_normalization_128 (BatchN (None, 17, 17, 128)  384
conv2d_128[0][0]
__________________________________________________________________________________
__________________
activation_128 (Activation)     (None, 17, 17, 128)  0
batch_normalization_128[0][0]
__________________________________________________________________________________
__________________
conv2d_129 (Conv2D)             (None, 17, 17, 128)  114688
activation_128[0][0]
__________________________________________________________________________________
__________________
batch_normalization_129 (BatchN (None, 17, 17, 128)  384
conv2d_129[0][0]
__________________________________________________________________________________
__________________
activation_129 (Activation)     (None, 17, 17, 128)  0
batch_normalization_129[0][0]
__________________________________________________________________________________
__________________
```

```
conv2d_125 (Conv2D)             (None, 17, 17, 128)  98304       mixed3[0][0]
------------------------------------------------------------------------------------------
------------------
conv2d_130 (Conv2D)             (None, 17, 17, 128)  114688
activation_129[0][0]
------------------------------------------------------------------------------------------
------------------
batch_normalization_125 (BatchN (None, 17, 17, 128)  384
conv2d_125[0][0]
------------------------------------------------------------------------------------------
------------------
batch_normalization_130 (BatchN (None, 17, 17, 128)  384
conv2d_130[0][0]
------------------------------------------------------------------------------------------
------------------
activation_125 (Activation)     (None, 17, 17, 128)  0
batch_normalization_125[0][0]
------------------------------------------------------------------------------------------
------------------
activation_130 (Activation)     (None, 17, 17, 128)  0
batch_normalization_130[0][0]
------------------------------------------------------------------------------------------
------------------
conv2d_126 (Conv2D)             (None, 17, 17, 128)  114688
activation_125[0][0]
------------------------------------------------------------------------------------------
------------------
conv2d_131 (Conv2D)             (None, 17, 17, 128)  114688
activation_130[0][0]
------------------------------------------------------------------------------------------
------------------
batch_normalization_126 (BatchN (None, 17, 17, 128)  384
conv2d_126[0][0]
------------------------------------------------------------------------------------------
------------------
batch_normalization_131 (BatchN (None, 17, 17, 128)  384
conv2d_131[0][0]
------------------------------------------------------------------------------------------
------------------
activation_126 (Activation)     (None, 17, 17, 128)  0
batch_normalization_126[0][0]
------------------------------------------------------------------------------------------
------------------
activation_131 (Activation)     (None, 17, 17, 128)  0
batch_normalization_131[0][0]
------------------------------------------------------------------------------------------
------------------
average_pooling2d_12 (AveragePo (None, 17, 17, 768)  0           mixed3[0][0]
```

```
__________________________________________________________________________________
__________________
conv2d_124 (Conv2D)             (None, 17, 17, 192)  147456      mixed3[0][0]
__________________________________________________________________________________
__________________
conv2d_127 (Conv2D)             (None, 17, 17, 192)  172032
activation_126[0][0]
__________________________________________________________________________________
__________________
conv2d_132 (Conv2D)             (None, 17, 17, 192)  172032
activation_131[0][0]
__________________________________________________________________________________
__________________
conv2d_133 (Conv2D)             (None, 17, 17, 192)  147456
average_pooling2d_12[0][0]
__________________________________________________________________________________
__________________
batch_normalization_124 (BatchN (None, 17, 17, 192)  576
conv2d_124[0][0]
__________________________________________________________________________________
__________________
batch_normalization_127 (BatchN (None, 17, 17, 192)  576
conv2d_127[0][0]
__________________________________________________________________________________
__________________
batch_normalization_132 (BatchN (None, 17, 17, 192)  576
conv2d_132[0][0]
__________________________________________________________________________________
__________________
batch_normalization_133 (BatchN (None, 17, 17, 192)  576
conv2d_133[0][0]
__________________________________________________________________________________
__________________
activation_124 (Activation)     (None, 17, 17, 192)  0
batch_normalization_124[0][0]
__________________________________________________________________________________
__________________
activation_127 (Activation)     (None, 17, 17, 192)  0
batch_normalization_127[0][0]
__________________________________________________________________________________
__________________
activation_132 (Activation)     (None, 17, 17, 192)  0
batch_normalization_132[0][0]
__________________________________________________________________________________
__________________
activation_133 (Activation)     (None, 17, 17, 192)  0
batch_normalization_133[0][0]
__________________________________________________________________________________
```

```
------------------
mixed4 (Concatenate)            (None, 17, 17, 768)  0
activation_124[0][0]
activation_127[0][0]
activation_132[0][0]
activation_133[0][0]
--------------------------------------------------------------------------------
------------------
conv2d_138 (Conv2D)             (None, 17, 17, 160)  122880      mixed4[0][0]
--------------------------------------------------------------------------------
------------------
batch_normalization_138 (BatchN (None, 17, 17, 160)  480
conv2d_138[0][0]
--------------------------------------------------------------------------------
------------------
activation_138 (Activation)     (None, 17, 17, 160)  0
batch_normalization_138[0][0]
--------------------------------------------------------------------------------
------------------
conv2d_139 (Conv2D)             (None, 17, 17, 160)  179200
activation_138[0][0]
--------------------------------------------------------------------------------
------------------
batch_normalization_139 (BatchN (None, 17, 17, 160)  480
conv2d_139[0][0]
--------------------------------------------------------------------------------
------------------
activation_139 (Activation)     (None, 17, 17, 160)  0
batch_normalization_139[0][0]
--------------------------------------------------------------------------------
------------------
conv2d_135 (Conv2D)             (None, 17, 17, 160)  122880      mixed4[0][0]
--------------------------------------------------------------------------------
------------------
conv2d_140 (Conv2D)             (None, 17, 17, 160)  179200
activation_139[0][0]
--------------------------------------------------------------------------------
------------------
batch_normalization_135 (BatchN (None, 17, 17, 160)  480
conv2d_135[0][0]
--------------------------------------------------------------------------------
------------------
batch_normalization_140 (BatchN (None, 17, 17, 160)  480
conv2d_140[0][0]
--------------------------------------------------------------------------------
------------------
activation_135 (Activation)     (None, 17, 17, 160)  0
batch_normalization_135[0][0]
```

```
____________________________________________________________________________
__________________
activation_140 (Activation)     (None, 17, 17, 160)  0
batch_normalization_140[0][0]
____________________________________________________________________________
__________________
conv2d_136 (Conv2D)             (None, 17, 17, 160)  179200
activation_135[0][0]
____________________________________________________________________________
__________________
conv2d_141 (Conv2D)             (None, 17, 17, 160)  179200
activation_140[0][0]
____________________________________________________________________________
__________________
batch_normalization_136 (BatchN (None, 17, 17, 160)  480
conv2d_136[0][0]
____________________________________________________________________________
__________________
batch_normalization_141 (BatchN (None, 17, 17, 160)  480
conv2d_141[0][0]
____________________________________________________________________________
__________________
activation_136 (Activation)     (None, 17, 17, 160)  0
batch_normalization_136[0][0]
____________________________________________________________________________
__________________
activation_141 (Activation)     (None, 17, 17, 160)  0
batch_normalization_141[0][0]
____________________________________________________________________________
__________________
average_pooling2d_13 (AveragePo (None, 17, 17, 768)  0           mixed4[0][0]
____________________________________________________________________________
__________________
conv2d_134 (Conv2D)             (None, 17, 17, 192)  147456      mixed4[0][0]
____________________________________________________________________________
__________________
conv2d_137 (Conv2D)             (None, 17, 17, 192)  215040
activation_136[0][0]
____________________________________________________________________________
__________________
conv2d_142 (Conv2D)             (None, 17, 17, 192)  215040
activation_141[0][0]
____________________________________________________________________________
__________________
conv2d_143 (Conv2D)             (None, 17, 17, 192)  147456
average_pooling2d_13[0][0]
____________________________________________________________________________
__________________
```

```
batch_normalization_134 (BatchN (None, 17, 17, 192)  576
conv2d_134[0][0]
__________________________________________________________________________________
__________________
batch_normalization_137 (BatchN (None, 17, 17, 192)  576
conv2d_137[0][0]
__________________________________________________________________________________
__________________
batch_normalization_142 (BatchN (None, 17, 17, 192)  576
conv2d_142[0][0]
__________________________________________________________________________________
__________________
batch_normalization_143 (BatchN (None, 17, 17, 192)  576
conv2d_143[0][0]
__________________________________________________________________________________
__________________
activation_134 (Activation)     (None, 17, 17, 192)  0
batch_normalization_134[0][0]
__________________________________________________________________________________
__________________
activation_137 (Activation)     (None, 17, 17, 192)  0
batch_normalization_137[0][0]
__________________________________________________________________________________
__________________
activation_142 (Activation)     (None, 17, 17, 192)  0
batch_normalization_142[0][0]
__________________________________________________________________________________
__________________
activation_143 (Activation)     (None, 17, 17, 192)  0
batch_normalization_143[0][0]
__________________________________________________________________________________
__________________
mixed5 (Concatenate)            (None, 17, 17, 768)  0
activation_134[0][0]
activation_137[0][0]
activation_142[0][0]
activation_143[0][0]
__________________________________________________________________________________
__________________
conv2d_148 (Conv2D)             (None, 17, 17, 160)  122880      mixed5[0][0]
__________________________________________________________________________________
__________________
batch_normalization_148 (BatchN (None, 17, 17, 160)  480
conv2d_148[0][0]
__________________________________________________________________________________
__________________
activation_148 (Activation)     (None, 17, 17, 160)  0
batch_normalization_148[0][0]
```

```
__________________________________________________________________________________
__________________
conv2d_149 (Conv2D)             (None, 17, 17, 160)  179200
activation_148[0][0]
__________________________________________________________________________________
__________________
batch_normalization_149 (BatchN (None, 17, 17, 160)  480
conv2d_149[0][0]
__________________________________________________________________________________
__________________
activation_149 (Activation)     (None, 17, 17, 160)  0
batch_normalization_149[0][0]
__________________________________________________________________________________
__________________
conv2d_145 (Conv2D)             (None, 17, 17, 160)  122880      mixed5[0][0]
__________________________________________________________________________________
__________________
conv2d_150 (Conv2D)             (None, 17, 17, 160)  179200
activation_149[0][0]
__________________________________________________________________________________
__________________
batch_normalization_145 (BatchN (None, 17, 17, 160)  480
conv2d_145[0][0]
__________________________________________________________________________________
__________________
batch_normalization_150 (BatchN (None, 17, 17, 160)  480
conv2d_150[0][0]
__________________________________________________________________________________
__________________
activation_145 (Activation)     (None, 17, 17, 160)  0
batch_normalization_145[0][0]
__________________________________________________________________________________
__________________
activation_150 (Activation)     (None, 17, 17, 160)  0
batch_normalization_150[0][0]
__________________________________________________________________________________
__________________
conv2d_146 (Conv2D)             (None, 17, 17, 160)  179200
activation_145[0][0]
__________________________________________________________________________________
__________________
conv2d_151 (Conv2D)             (None, 17, 17, 160)  179200
activation_150[0][0]
__________________________________________________________________________________
__________________
batch_normalization_146 (BatchN (None, 17, 17, 160)  480
conv2d_146[0][0]
__________________________________________________________________________________
```

```
__________________
batch_normalization_151 (BatchN (None, 17, 17, 160)  480
conv2d_151[0][0]
__________________________________________________________________________________
__________________
activation_146 (Activation)     (None, 17, 17, 160)  0
batch_normalization_146[0][0]
__________________________________________________________________________________
__________________
activation_151 (Activation)     (None, 17, 17, 160)  0
batch_normalization_151[0][0]
__________________________________________________________________________________
__________________
average_pooling2d_14 (AveragePo (None, 17, 17, 768)  0           mixed5[0][0]
__________________________________________________________________________________
__________________
conv2d_144 (Conv2D)             (None, 17, 17, 192)  147456      mixed5[0][0]
__________________________________________________________________________________
__________________
conv2d_147 (Conv2D)             (None, 17, 17, 192)  215040
activation_146[0][0]
__________________________________________________________________________________
__________________
conv2d_152 (Conv2D)             (None, 17, 17, 192)  215040
activation_151[0][0]
__________________________________________________________________________________
__________________
conv2d_153 (Conv2D)             (None, 17, 17, 192)  147456
average_pooling2d_14[0][0]
__________________________________________________________________________________
__________________
batch_normalization_144 (BatchN (None, 17, 17, 192)  576
conv2d_144[0][0]
__________________________________________________________________________________
__________________
batch_normalization_147 (BatchN (None, 17, 17, 192)  576
conv2d_147[0][0]
__________________________________________________________________________________
__________________
batch_normalization_152 (BatchN (None, 17, 17, 192)  576
conv2d_152[0][0]
__________________________________________________________________________________
__________________
batch_normalization_153 (BatchN (None, 17, 17, 192)  576
conv2d_153[0][0]
__________________________________________________________________________________
__________________
activation_144 (Activation)     (None, 17, 17, 192)  0
```

```
batch_normalization_144[0][0]
__________________________________________________________________________________
__________________
activation_147 (Activation)     (None, 17, 17, 192)  0
batch_normalization_147[0][0]
__________________________________________________________________________________
__________________
activation_152 (Activation)     (None, 17, 17, 192)  0
batch_normalization_152[0][0]
__________________________________________________________________________________
__________________
activation_153 (Activation)     (None, 17, 17, 192)  0
batch_normalization_153[0][0]
__________________________________________________________________________________
__________________
mixed6 (Concatenate)            (None, 17, 17, 768)  0
activation_144[0][0]
activation_147[0][0]
activation_152[0][0]
activation_153[0][0]
__________________________________________________________________________________
__________________
conv2d_158 (Conv2D)             (None, 17, 17, 192)  147456      mixed6[0][0]
__________________________________________________________________________________
__________________
batch_normalization_158 (BatchN (None, 17, 17, 192)  576
conv2d_158[0][0]
__________________________________________________________________________________
__________________
activation_158 (Activation)     (None, 17, 17, 192)  0
batch_normalization_158[0][0]
__________________________________________________________________________________
__________________
conv2d_159 (Conv2D)             (None, 17, 17, 192)  258048
activation_158[0][0]
__________________________________________________________________________________
__________________
batch_normalization_159 (BatchN (None, 17, 17, 192)  576
conv2d_159[0][0]
__________________________________________________________________________________
__________________
activation_159 (Activation)     (None, 17, 17, 192)  0
batch_normalization_159[0][0]
__________________________________________________________________________________
__________________
conv2d_155 (Conv2D)             (None, 17, 17, 192)  147456      mixed6[0][0]
__________________________________________________________________________________
__________________
```

```
conv2d_160 (Conv2D)             (None, 17, 17, 192)  258048
activation_159[0][0]
__________________________________________________________________________________
__________________
batch_normalization_155 (BatchN (None, 17, 17, 192)  576
conv2d_155[0][0]
__________________________________________________________________________________
__________________
batch_normalization_160 (BatchN (None, 17, 17, 192)  576
conv2d_160[0][0]
__________________________________________________________________________________
__________________
activation_155 (Activation)     (None, 17, 17, 192)  0
batch_normalization_155[0][0]
__________________________________________________________________________________
__________________
activation_160 (Activation)     (None, 17, 17, 192)  0
batch_normalization_160[0][0]
__________________________________________________________________________________
__________________
conv2d_156 (Conv2D)             (None, 17, 17, 192)  258048
activation_155[0][0]
__________________________________________________________________________________
__________________
conv2d_161 (Conv2D)             (None, 17, 17, 192)  258048
activation_160[0][0]
__________________________________________________________________________________
__________________
batch_normalization_156 (BatchN (None, 17, 17, 192)  576
conv2d_156[0][0]
__________________________________________________________________________________
__________________
batch_normalization_161 (BatchN (None, 17, 17, 192)  576
conv2d_161[0][0]
__________________________________________________________________________________
__________________
activation_156 (Activation)     (None, 17, 17, 192)  0
batch_normalization_156[0][0]
__________________________________________________________________________________
__________________
activation_161 (Activation)     (None, 17, 17, 192)  0
batch_normalization_161[0][0]
__________________________________________________________________________________
__________________
average_pooling2d_15 (AveragePo (None, 17, 17, 768)  0           mixed6[0][0]
__________________________________________________________________________________
__________________
conv2d_154 (Conv2D)             (None, 17, 17, 192)  147456      mixed6[0][0]
```

```
__________________________________________________________________________________
__________________
conv2d_157 (Conv2D)             (None, 17, 17, 192)  258048
activation_156[0][0]
__________________________________________________________________________________
__________________
conv2d_162 (Conv2D)             (None, 17, 17, 192)  258048
activation_161[0][0]
__________________________________________________________________________________
__________________
conv2d_163 (Conv2D)             (None, 17, 17, 192)  147456
average_pooling2d_15[0][0]
__________________________________________________________________________________
__________________
batch_normalization_154 (BatchN (None, 17, 17, 192)  576
conv2d_154[0][0]
__________________________________________________________________________________
__________________
batch_normalization_157 (BatchN (None, 17, 17, 192)  576
conv2d_157[0][0]
__________________________________________________________________________________
__________________
batch_normalization_162 (BatchN (None, 17, 17, 192)  576
conv2d_162[0][0]
__________________________________________________________________________________
__________________
batch_normalization_163 (BatchN (None, 17, 17, 192)  576
conv2d_163[0][0]
__________________________________________________________________________________
__________________
activation_154 (Activation)     (None, 17, 17, 192)  0
batch_normalization_154[0][0]
__________________________________________________________________________________
__________________
activation_157 (Activation)     (None, 17, 17, 192)  0
batch_normalization_157[0][0]
__________________________________________________________________________________
__________________
activation_162 (Activation)     (None, 17, 17, 192)  0
batch_normalization_162[0][0]
__________________________________________________________________________________
__________________
activation_163 (Activation)     (None, 17, 17, 192)  0
batch_normalization_163[0][0]
__________________________________________________________________________________
__________________
mixed7 (Concatenate)            (None, 17, 17, 768)  0
activation_154[0][0]
```

```
activation_157[0][0]
activation_162[0][0]
activation_163[0][0]
__________________________________________________________________________________
__________________
conv2d_166 (Conv2D)             (None, 17, 17, 192)  147456      mixed7[0][0]
__________________________________________________________________________________
__________________
batch_normalization_166 (BatchN (None, 17, 17, 192)  576
conv2d_166[0][0]
__________________________________________________________________________________
__________________
activation_166 (Activation)     (None, 17, 17, 192)  0
batch_normalization_166[0][0]
__________________________________________________________________________________
__________________
conv2d_167 (Conv2D)             (None, 17, 17, 192)  258048
activation_166[0][0]
__________________________________________________________________________________
__________________
batch_normalization_167 (BatchN (None, 17, 17, 192)  576
conv2d_167[0][0]
__________________________________________________________________________________
__________________
activation_167 (Activation)     (None, 17, 17, 192)  0
batch_normalization_167[0][0]
__________________________________________________________________________________
__________________
conv2d_164 (Conv2D)             (None, 17, 17, 192)  147456      mixed7[0][0]
__________________________________________________________________________________
__________________
conv2d_168 (Conv2D)             (None, 17, 17, 192)  258048
activation_167[0][0]
__________________________________________________________________________________
__________________
batch_normalization_164 (BatchN (None, 17, 17, 192)  576
conv2d_164[0][0]
__________________________________________________________________________________
__________________
batch_normalization_168 (BatchN (None, 17, 17, 192)  576
conv2d_168[0][0]
__________________________________________________________________________________
__________________
activation_164 (Activation)     (None, 17, 17, 192)  0
batch_normalization_164[0][0]
__________________________________________________________________________________
__________________
activation_168 (Activation)     (None, 17, 17, 192)  0
```

```
batch_normalization_168[0][0]
__________________________________________________________________________________________________
conv2d_165 (Conv2D)             (None, 8, 8, 320)    552960
activation_164[0][0]
__________________________________________________________________________________________________
conv2d_169 (Conv2D)             (None, 8, 8, 192)    331776
activation_168[0][0]
__________________________________________________________________________________________________
batch_normalization_165 (BatchN (None, 8, 8, 320)    960
conv2d_165[0][0]
__________________________________________________________________________________________________
batch_normalization_169 (BatchN (None, 8, 8, 192)    576
conv2d_169[0][0]
__________________________________________________________________________________________________
activation_165 (Activation)     (None, 8, 8, 320)    0
batch_normalization_165[0][0]
__________________________________________________________________________________________________
activation_169 (Activation)     (None, 8, 8, 192)    0
batch_normalization_169[0][0]
__________________________________________________________________________________________________
max_pooling2d_7 (MaxPooling2D)  (None, 8, 8, 768)    0           mixed7[0][0]
__________________________________________________________________________________________________
mixed8 (Concatenate)            (None, 8, 8, 1280)   0
activation_165[0][0]
activation_169[0][0]
max_pooling2d_7[0][0]
__________________________________________________________________________________________________
conv2d_174 (Conv2D)             (None, 8, 8, 448)    573440      mixed8[0][0]
__________________________________________________________________________________________________
batch_normalization_174 (BatchN (None, 8, 8, 448)    1344
conv2d_174[0][0]
__________________________________________________________________________________________________
activation_174 (Activation)     (None, 8, 8, 448)    0
batch_normalization_174[0][0]
__________________________________________________________________________________________________
conv2d_171 (Conv2D)             (None, 8, 8, 384)    491520      mixed8[0][0]
```

```
__________________________________________________________________________________
__________________
conv2d_175 (Conv2D)             (None, 8, 8, 384)    1548288
activation_174[0][0]
__________________________________________________________________________________
__________________
batch_normalization_171 (BatchN (None, 8, 8, 384)    1152
conv2d_171[0][0]
__________________________________________________________________________________
__________________
batch_normalization_175 (BatchN (None, 8, 8, 384)    1152
conv2d_175[0][0]
__________________________________________________________________________________
__________________
activation_171 (Activation)     (None, 8, 8, 384)    0
batch_normalization_171[0][0]
__________________________________________________________________________________
__________________
activation_175 (Activation)     (None, 8, 8, 384)    0
batch_normalization_175[0][0]
__________________________________________________________________________________
__________________
conv2d_172 (Conv2D)             (None, 8, 8, 384)    442368
activation_171[0][0]
__________________________________________________________________________________
__________________
conv2d_173 (Conv2D)             (None, 8, 8, 384)    442368
activation_171[0][0]
__________________________________________________________________________________
__________________
conv2d_176 (Conv2D)             (None, 8, 8, 384)    442368
activation_175[0][0]
__________________________________________________________________________________
__________________
conv2d_177 (Conv2D)             (None, 8, 8, 384)    442368
activation_175[0][0]
__________________________________________________________________________________
__________________
average_pooling2d_16 (AveragePo (None, 8, 8, 1280)   0           mixed8[0][0]
__________________________________________________________________________________
__________________
conv2d_170 (Conv2D)             (None, 8, 8, 320)    409600      mixed8[0][0]
__________________________________________________________________________________
__________________
batch_normalization_172 (BatchN (None, 8, 8, 384)    1152
conv2d_172[0][0]
__________________________________________________________________________________
__________________
```

```
batch_normalization_173 (BatchN (None, 8, 8, 384)    1152
conv2d_173[0][0]
__________________________________________________________________________________
__________________
batch_normalization_176 (BatchN (None, 8, 8, 384)    1152
conv2d_176[0][0]
__________________________________________________________________________________
__________________
batch_normalization_177 (BatchN (None, 8, 8, 384)    1152
conv2d_177[0][0]
__________________________________________________________________________________
__________________
conv2d_178 (Conv2D)             (None, 8, 8, 192)    245760
average_pooling2d_16[0][0]
__________________________________________________________________________________
__________________
batch_normalization_170 (BatchN (None, 8, 8, 320)    960
conv2d_170[0][0]
__________________________________________________________________________________
__________________
activation_172 (Activation)     (None, 8, 8, 384)    0
batch_normalization_172[0][0]
__________________________________________________________________________________
__________________
activation_173 (Activation)     (None, 8, 8, 384)    0
batch_normalization_173[0][0]
__________________________________________________________________________________
__________________
activation_176 (Activation)     (None, 8, 8, 384)    0
batch_normalization_176[0][0]
__________________________________________________________________________________
__________________
activation_177 (Activation)     (None, 8, 8, 384)    0
batch_normalization_177[0][0]
__________________________________________________________________________________
__________________
batch_normalization_178 (BatchN (None, 8, 8, 192)    576
conv2d_178[0][0]
__________________________________________________________________________________
__________________
activation_170 (Activation)     (None, 8, 8, 320)    0
batch_normalization_170[0][0]
__________________________________________________________________________________
__________________
mixed9_0 (Concatenate)          (None, 8, 8, 768)    0
activation_172[0][0]
activation_173[0][0]
__________________________________________________________________________________
```

```
------------------
concatenate_2 (Concatenate)     (None, 8, 8, 768)    0
activation_176[0][0]
activation_177[0][0]
--------------------------------------------------------------------------------
------------------
activation_178 (Activation)     (None, 8, 8, 192)    0
batch_normalization_178[0][0]
--------------------------------------------------------------------------------
------------------
mixed9 (Concatenate)            (None, 8, 8, 2048)   0
activation_170[0][0]
                                                                  mixed9_0[0][0]
concatenate_2[0][0]
activation_178[0][0]
--------------------------------------------------------------------------------
------------------
conv2d_183 (Conv2D)             (None, 8, 8, 448)    917504      mixed9[0][0]
--------------------------------------------------------------------------------
------------------
batch_normalization_183 (BatchN (None, 8, 8, 448)    1344
conv2d_183[0][0]
--------------------------------------------------------------------------------
------------------
activation_183 (Activation)     (None, 8, 8, 448)    0
batch_normalization_183[0][0]
--------------------------------------------------------------------------------
------------------
conv2d_180 (Conv2D)             (None, 8, 8, 384)    786432      mixed9[0][0]
--------------------------------------------------------------------------------
------------------
conv2d_184 (Conv2D)             (None, 8, 8, 384)    1548288
activation_183[0][0]
--------------------------------------------------------------------------------
------------------
batch_normalization_180 (BatchN (None, 8, 8, 384)    1152
conv2d_180[0][0]
--------------------------------------------------------------------------------
------------------
batch_normalization_184 (BatchN (None, 8, 8, 384)    1152
conv2d_184[0][0]
--------------------------------------------------------------------------------
------------------
activation_180 (Activation)     (None, 8, 8, 384)    0
batch_normalization_180[0][0]
--------------------------------------------------------------------------------
------------------
activation_184 (Activation)     (None, 8, 8, 384)    0
```

```
batch_normalization_184[0][0]
__________________________________________________________________________________________________
conv2d_181 (Conv2D)             (None, 8, 8, 384)    442368
activation_180[0][0]
__________________________________________________________________________________________________
conv2d_182 (Conv2D)             (None, 8, 8, 384)    442368
activation_180[0][0]
__________________________________________________________________________________________________
conv2d_185 (Conv2D)             (None, 8, 8, 384)    442368
activation_184[0][0]
__________________________________________________________________________________________________
conv2d_186 (Conv2D)             (None, 8, 8, 384)    442368
activation_184[0][0]
__________________________________________________________________________________________________
average_pooling2d_17 (AveragePo (None, 8, 8, 2048)   0           mixed9[0][0]
__________________________________________________________________________________________________
conv2d_179 (Conv2D)             (None, 8, 8, 320)    655360      mixed9[0][0]
__________________________________________________________________________________________________
batch_normalization_181 (BatchN (None, 8, 8, 384)    1152
conv2d_181[0][0]
__________________________________________________________________________________________________
batch_normalization_182 (BatchN (None, 8, 8, 384)    1152
conv2d_182[0][0]
__________________________________________________________________________________________________
batch_normalization_185 (BatchN (None, 8, 8, 384)    1152
conv2d_185[0][0]
__________________________________________________________________________________________________
batch_normalization_186 (BatchN (None, 8, 8, 384)    1152
conv2d_186[0][0]
__________________________________________________________________________________________________
conv2d_187 (Conv2D)             (None, 8, 8, 192)    393216
average_pooling2d_17[0][0]
__________________________________________________________________________________________________
batch_normalization_179 (BatchN (None, 8, 8, 320)    960
conv2d_179[0][0]
__________________________________________________________________________________________________
```

```
__________________
activation_181 (Activation)     (None, 8, 8, 384)    0
batch_normalization_181[0][0]

__________________________________________________________________________________
__________________
activation_182 (Activation)     (None, 8, 8, 384)    0
batch_normalization_182[0][0]

__________________________________________________________________________________
__________________
activation_185 (Activation)     (None, 8, 8, 384)    0
batch_normalization_185[0][0]

__________________________________________________________________________________
__________________
activation_186 (Activation)     (None, 8, 8, 384)    0
batch_normalization_186[0][0]

__________________________________________________________________________________
__________________
batch_normalization_187 (BatchN (None, 8, 8, 192)    576
conv2d_187[0][0]

__________________________________________________________________________________
__________________
activation_179 (Activation)     (None, 8, 8, 320)    0
batch_normalization_179[0][0]

__________________________________________________________________________________
__________________
mixed9_1 (Concatenate)          (None, 8, 8, 768)    0
activation_181[0][0]
activation_182[0][0]

__________________________________________________________________________________
__________________
concatenate_3 (Concatenate)     (None, 8, 8, 768)    0
activation_185[0][0]
activation_186[0][0]

__________________________________________________________________________________
__________________
activation_187 (Activation)     (None, 8, 8, 192)    0
batch_normalization_187[0][0]

__________________________________________________________________________________
__________________
mixed10 (Concatenate)           (None, 8, 8, 2048)   0
activation_179[0][0]
                                                                 mixed9_1[0][0]
concatenate_3[0][0]
activation_187[0][0]

__________________________________________________________________________________
__________________
avg_pool (GlobalAveragePooling2 (None, 2048)         0           mixed10[0][0]
==================================================================================
```

```
==================
Total params: 21,802,784
Trainable params: 21,768,352
Non-trainable params: 34,432
_________________________________________________________________________________
_________________
```

[98]:
```
# Performing image embedding with featExtractor
def encodeData(curr_image,featExtractor):
    curr_image = preprocessingUtil(curr_image)
    featVector = featExtractor.predict(curr_image)
    featVector = np.ravel(featVector)
    return featVector
```

[108]:
```
# Dividing the data into training, validation, and testing data
AllData = list(my_images)
trainingData = AllData[:6000]
validationData = AllData[6000:7000]
TestingData = AllData[7000:]
print(f'Number of training images {len(trainingData)}')
print(f'Number of testing images {len(TestingData)}')
print(f'Number of validation images {len(validationData)}')
```

```
Number of training images 6000
Number of testing images 1091
Number of validation images 1000
```

[109]:
```
trainingEnc = {}
validationEnc = {}
testingEnc = {}
```

[110]:
```
# Encoding the training data
for x in range(1,len(trainingData)):
    myloc = '../input/flickr8k/Images/' + trainingData[x]
    trainingEnc[trainingData[x]] = encodeData(myloc,featExtractor)
```

[111]:
```
# Storing the encoded training images
with open('./training_images_enc.pkl','wb') as pickEnc:
    dump(trainingEnc,pickEnc)
```

[112]:
```
# Similarly we encode and store the validation and testing data
for curr_img in tqdm(validationData):
    validationEnc[curr_img] = encodeData('../input/flickr8k/Images/' +␣
 ↪curr_img,featExtractor)
```

```
100%|      | 1000/1000 [01:07<00:00, 14.88it/s]
```

[113]:
```
with open('./validation_images_enc.pkl','wb') as pickEnc:
    dump(validationEnc,pickEnc)
```

[114]:
```
for curr_img in tqdm(TestingData):
    testingEnc[curr_img] = encodeData('../input/flickr8k/Images/' +␣
 ↪curr_img,featExtractor)
```

```
100%|      | 1091/1091 [01:12<00:00, 15.00it/s]
```

[115]:
```
with open('./testing_images_enc.pkl','wb') as pickEnc:
    dump(testingEnc,pickEnc)
```

[116]:
```
!cp ./testing_images_enc.pkl /content/drive/MyDrive
```

```
cp: cannot create regular file '/content/drive/MyDrive': No such file or
directory
```

[117]:
```
# Loading the features
featTrain = load(open('./training_images_enc.pkl','rb'))

featValid = load(open('./validation_images_enc.pkl','rb'))

featTest = load(open('./testing_images_enc.pkl','rb'))
print(f'The number of training features {len(featTrain)}')
print(f'The number of testing features {len(featTest)}')
print(f'The number of validation features {len(featValid)}')
```

```
The number of training features 5999
The number of testing features 1091
The number of validation features 1000
```

[118]:
```
# Creating separate lists for storing the training, testing, and validation␣
 ↪captions
allCaptionsTrain = []
allCaptionsVal = []
allCaptionsTest = []
for x in range(1,len(trainingData)):
    curr_img=trainingData[x]
    captions = imgNameDict[curr_img]
    for mycap in captions:
        allCaptionsTrain.append(mycap)
print(f'Number of all training captions {len(allCaptionsTrain)}')
```

```
Number of all training captions 29995
```

[119]:
```
for curr_img in validationData:
    captions = imgNameDict[curr_img]
    for mycap in captions:
        allCaptionsVal.append(mycap)
print(f'Number of all validation captions {len(allCaptionsVal)}')
```

```
Number of all validation captions 5000
```

[120]:
```
for img in TestingData:
    captions = imgNameDict[img]
    for mycap in captions:
        allCaptionsTest.append(mycap)
print(f'Number of all test captions {len(allCaptionsTest)}')
```

```
Number of all test captions 5455
```

[121]:
```
wc = {}
for mysentence in allCaptionsTrain:
    for curr_word in mysentence.split(' '):
        wc[curr_word] = wc.get(curr_word,0) + 1

myvocab = [curr_word for curr_word in wc if wc[curr_word] >= 10]
```

[122]:
```
id_word = {}
word_id = {}
idx = 1
for curr_word in myvocab:
    word_id[curr_word] = idx
    id_word[idx] = curr_word
    idx +=1
```

[123]:
```
with open('./word_id.pkl','wb') as word_id1:
    dump(word_id,word_id1)
```

[124]:
```
with open('./id_word.pkl','wb') as id_word1:
    dump(id_word,id_word1)
```

[125]:
```
vocabSize = len(id_word) +1
vocabSize
```

[125]:
```
1621
```

[126]:
```
trainDict = {}
for x in range(1,len(trainingData)):
    my_image=trainingData[x]
    trainDict[my_image] = imgNameDict[my_image]
```

[127]:
```
# Defining function to find maximum length of captions
def maxFun(imageCaptions):
    lines = to_lines(imageCaptions)
    return max(len(line.split()) for line in lines)

# Defining function which accepts dictionary of clean captions as input
# Returns captions list
def to_lines(imageCaptions):
    allCaptions = []
```

```
    for x in imageCaptions.keys():
        [allCaptions.append(mycap) for mycap in imageCaptions[x]]
    return allCaptions

maxLength = maxFun(trainDict)
print(f'The captions maximum length is {maxLength}')
```

```
The captions maximum length is 30
```

[128]:
```
# Loading GloVe
embeddingsIndex = {}
myfile = open('../input/glove6b/glove.6B.200d.txt',encoding = 'utf-8')

for x in tqdm(myfile):
    vals = x.split()
    curr_word = vals[0]
    coefficients = np.asarray(vals[1:],dtype = 'float32')
    embeddingsIndex[curr_word] = coefficients

myfile.close()
print('Number of word vectors is %s' % len(embeddingsIndex))
```

```
400000it [00:18, 21609.09it/s]

Number of word vectors is 400000
```

[129]:
```
embedDim = 200
```

[130]:
```
# Creating the embedding matrix
embedMat = np.zeros((vocabSize,embedDim))
print(f'The embedding matrix dimensions are {embedMat.shape}')
```

```
The embedding matrix dimensions are (1621, 200)
```

[131]:
```
for curr_word,idx in word_id.items():
    embedVec = embeddingsIndex.get(curr_word)
    if embedVec is not None:
        embedMat[idx] = embedVec
```

[132]:
```
# Building the model
inp1 = Input(shape = (2048,))
fe1 = Dropout(0.5)(inp1)
fe2 = Dense(256,activation = 'relu')(fe1)
inp2 = Input(shape = (maxLength,))
se1 = Embedding(vocabSize,embedDim,mask_zero = True)(inp2)
se2 = Dropout(0.5)(se1)
se3 = LSTM(256)(se2)
dec1 = add([fe2,se3])
```

```
dec2 = Dense(256,activation = 'relu')(dec1)
out = Dense(vocabSize,activation = 'softmax')(dec2)
```

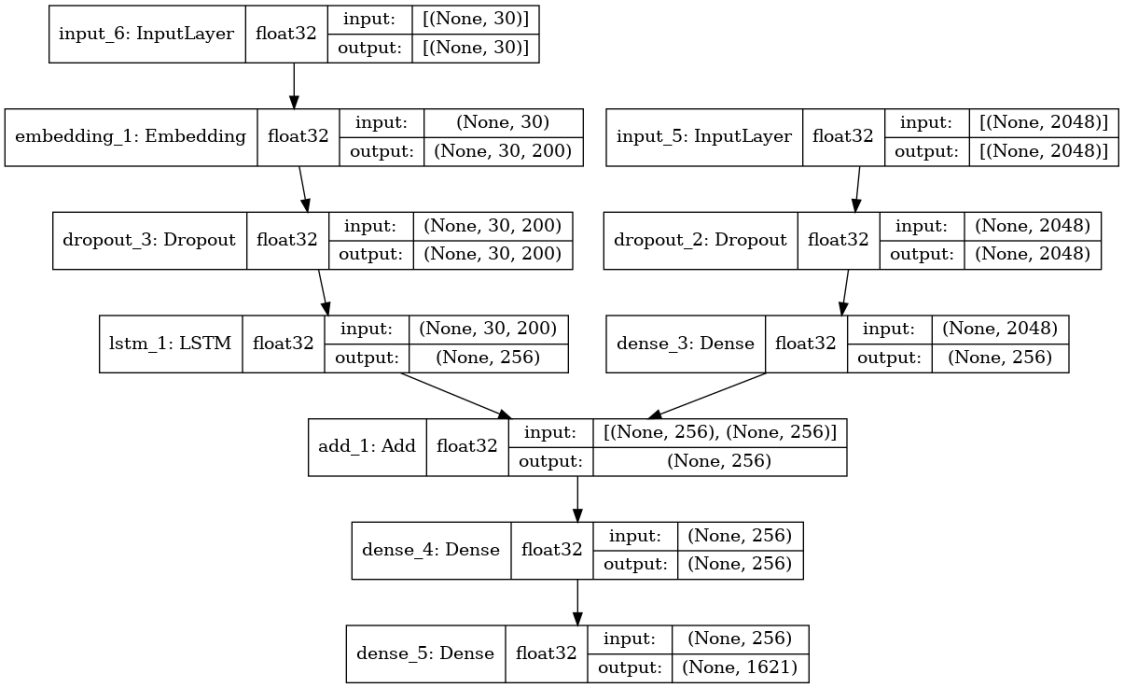
[133]:
```
mymodel2 = Model(inputs = [inp1,inp2],outputs = out)
mymodel2.summary()
```

```
Model: "model_3"
__________________________________________________________________________________________________
Layer (type)                    Output Shape         Param #     Connected to
==================================================================================================
input_6 (InputLayer)            [(None, 30)]         0
__________________________________________________________________________________________________
input_5 (InputLayer)            [(None, 2048)]       0
__________________________________________________________________________________________________
embedding_1 (Embedding)         (None, 30, 200)      324200      input_6[0][0]
__________________________________________________________________________________________________
dropout_2 (Dropout)             (None, 2048)         0           input_5[0][0]
__________________________________________________________________________________________________
dropout_3 (Dropout)             (None, 30, 200)      0
embedding_1[0][0]
__________________________________________________________________________________________________
dense_3 (Dense)                 (None, 256)          524544      dropout_2[0][0]
__________________________________________________________________________________________________
lstm_1 (LSTM)                   (None, 256)          467968      dropout_3[0][0]
__________________________________________________________________________________________________
add_1 (Add)                     (None, 256)          0           dense_3[0][0]
                                                                 lstm_1[0][0]
__________________________________________________________________________________________________
dense_4 (Dense)                 (None, 256)          65792       add_1[0][0]
__________________________________________________________________________________________________
dense_5 (Dense)                 (None, 1621)         416597      dense_4[0][0]
==================================================================================================
Total params: 1,799,101
Trainable params: 1,799,101
Non-trainable params: 0
__________________________________________________________________________________________________
```

```
------------------
```

[134]:
```
# Displaying the model layers
plot_model(mymodel2,
           to_file='mymodel2.png',
           show_layer_names=True,
           show_dtype=True,
           show_shapes=True
           )
```

[134]:



[135]:
```
print(f'Embedding Matrix Shape {embedMat.shape}')
embedWeight = mymodel2.layers[2].get_weights()
print(f'Embedding Weight Shape {embedWeight[0].shape}')
mymodel2.layers[2].set_weights([embedMat])
mymodel2.layers[2].trainable = False
```

```
Embedding Matrix Shape (1621, 200)
Embedding Weight Shape (1621, 200)
```

[136]:
```
mymodel2.summary()
```

```
Model: "model_3"
__________________________________________________________________________________
__________________
Layer (type)                    Output Shape         Param #     Connected to
==================================================================================
==================
```

```
input_6 (InputLayer)            [(None, 30)]         0
__________________________________________________________________________________________
__________________
input_5 (InputLayer)            [(None, 2048)]       0
__________________________________________________________________________________________
__________________
embedding_1 (Embedding)         (None, 30, 200)      324200      input_6[0][0]
__________________________________________________________________________________________
__________________
dropout_2 (Dropout)             (None, 2048)         0           input_5[0][0]
__________________________________________________________________________________________
__________________
dropout_3 (Dropout)             (None, 30, 200)      0
embedding_1[0][0]
__________________________________________________________________________________________
__________________
dense_3 (Dense)                 (None, 256)          524544      dropout_2[0][0]
__________________________________________________________________________________________
__________________
lstm_1 (LSTM)                   (None, 256)          467968      dropout_3[0][0]
__________________________________________________________________________________________
__________________
add_1 (Add)                     (None, 256)          0           dense_3[0][0]
                                                                 lstm_1[0][0]
__________________________________________________________________________________________
__________________
dense_4 (Dense)                 (None, 256)          65792       add_1[0][0]
__________________________________________________________________________________________
__________________
dense_5 (Dense)                 (None, 1621)         416597      dense_4[0][0]
==========================================================================================
==================
Total params: 1,799,101
Trainable params: 1,474,901
Non-trainable params: 324,200
__________________________________________________________________________________________
__________________
```

[137]:
```
mymodel2.compile(optimizer = 'adam',loss = 'categorical_crossentropy')
```

[138]:
```
epochs = 10
steps = len(trainDict)//6

mymodel2.optimizer.lr = 1e-4
```

[ ]:
```

```

[139]:
```
def data_generator(trainDict,featTrain,word_id,maxLength,pics_p_batch):
    y=[]
    n = 0
    X1 =[]
    X2=[]

    while True:
        for curr_image,captionsList in trainDict.items():
            n+=1
            myfeatVect = featTrain[curr_image]
            for mycap in captionsList:
                encoded_cap = [word_id[word] for word in mycap.split(' ') if␣
 ↪word in word_id]
                for idx in range(1,len(encoded_cap)):
                    in_seq = encoded_cap[:idx]
                    out_seq = encoded_cap[idx]
                    in_seq = pad_sequences([in_seq],maxlen = maxLength)[0]
                    out_seq = to_categorical([out_seq],num_classes =␣
 ↪vocabSize)[0]
                    y.append(out_seq)
                    X1.append(myfeatVect)
                    X2.append(in_seq)

            if n == pics_p_batch:
                X1,X2,y = np.array(X1),np.array(X2),np.array(y)
                yield [X1,X2],y
                y=[]
                n=0
                X1 =[]
                X2=[]
```

[140]:
```
'3448855727_f16dea7b03.jpg' in featTrain
```

[140]: True

[141]:
```
for curr_epoch in range(epochs):
    genvar = data_generator(trainDict, featTrain, word_id, maxLength, 6)
    mymodel2.fit_generator(genvar, epochs=1, steps_per_epoch=steps, verbose=1)
```

```
/opt/conda/lib/python3.7/site-
packages/tensorflow/python/keras/engine/training.py:1844: UserWarning:
`Model.fit_generator` is deprecated and will be removed in a future version.
Please use `Model.fit`, which supports generators.
  warnings.warn('`Model.fit_generator` is deprecated and '

999/999 [==============================] - 113s 109ms/step - loss: 5.5713
999/999 [==============================] - 111s 111ms/step - loss: 4.4140
```

```
999/999 [==============================] - 117s 117ms/step - loss: 4.0736
999/999 [==============================] - 108s 108ms/step - loss: 3.8797
999/999 [==============================] - 110s 110ms/step - loss: 3.7496
999/999 [==============================] - 117s 118ms/step - loss: 3.6473
999/999 [==============================] - 108s 108ms/step - loss: 3.5669
999/999 [==============================] - 107s 107ms/step - loss: 3.4978
999/999 [==============================] - 117s 117ms/step - loss: 3.4407
999/999 [==============================] - 115s 115ms/step - loss: 3.3863
```

[142]:
```
# Saving the model
mymodel2.save('model_flickr8k')

# Saving the model weights
mymodel2.save_weights('./model_flickr8k.h5')
```

[143]:
```
!zip -r model_flickr8k.zip model_flickr8k
```

```
updating: model_flickr8k/ (stored 0%)
updating: model_flickr8k/saved_model.pb (deflated 90%)
updating: model_flickr8k/variables/ (stored 0%)
updating: model_flickr8k/variables/variables.data-00000-of-00001 (deflated 7%)
updating: model_flickr8k/variables/variables.index (deflated 64%)
updating: model_flickr8k/assets/ (stored 0%)
```

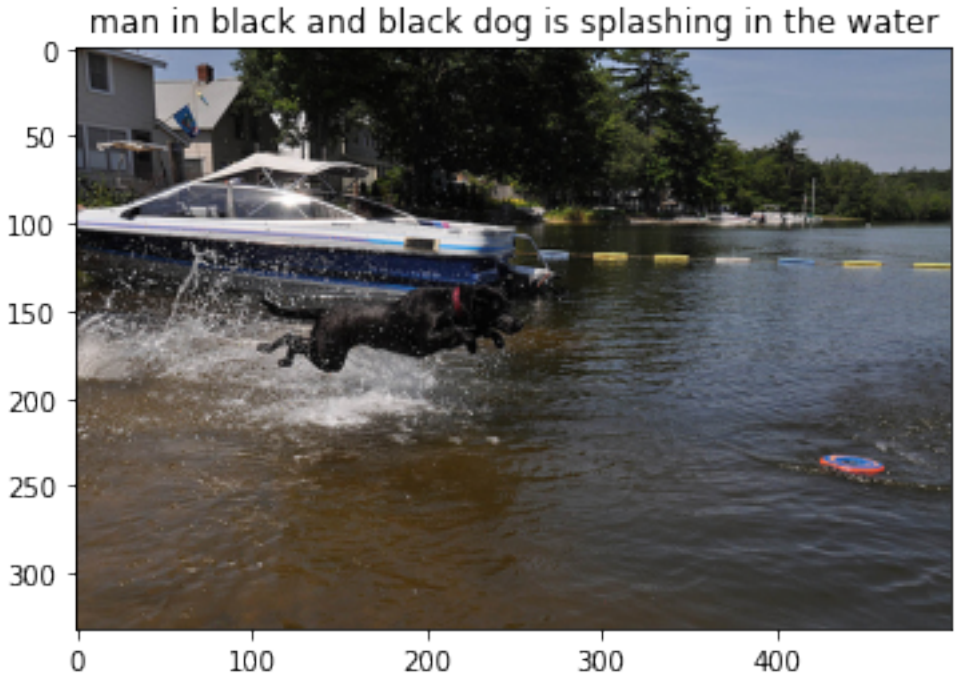
[144]:
```
#To obtain the predicted image caption, the image is passed to the model with␣
 ↪the input string "startseq".
# Then the model predicts the next word and this word is appended to the input␣
 ↪string. This repeats until
# "endseq" is reached or the maximum sentence length is reached.

def greedy_search(featVect,verbose = 0):
    in_text ='startseq'
    for i in range(maxLength):
        sequence = [word_id[x] for x in in_text.split() if x in word_id]
        sequence = pad_sequences([sequence],maxlen = maxLength)
        yhat = mymodel2.predict([featVect,sequence],verbose = verbose) #␣
 ↪[(1,2048),(1,31)]
        yhat = np.argmax(yhat)
        curr_word = id_word[yhat]
        in_text += ' ' + curr_word
        if curr_word == 'endseq':
            break
    res = in_text.split()
    res = res[1:-1]
    res = ' '.join(res)
    return res
```

```
img = random.choice(TestingData)
featVect = featTest[img].reshape((1,2048))
print(f'feature_vector: {featVect.shape}')
final = greedy_search(featVect,1)
final
```

```
feature_vector: (1, 2048)
1/1 [==============================] - 1s 1s/step
1/1 [==============================] - 0s 21ms/step
1/1 [==============================] - 0s 20ms/step
1/1 [==============================] - 0s 20ms/step
1/1 [==============================] - 0s 21ms/step
1/1 [==============================] - 0s 22ms/step
1/1 [==============================] - 0s 20ms/step
1/1 [==============================] - 0s 22ms/step
1/1 [==============================] - 0s 23ms/step
```

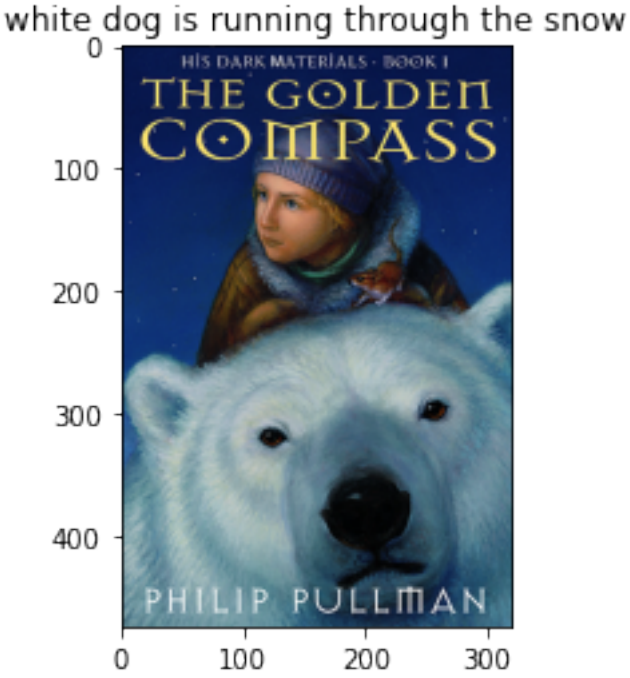
[144]:
```
'woman wearing red shirt and glasses and glasses'
```

[145]:
```
img = random.choice(TestingData)
feature_vector = featTest[img].reshape((1,2048))
x = plt.imread('../input/flickr8k/Images/' + img)
plt.imshow(x)
plt.title(greedy_search(feature_vector,0))
plt.show()
```



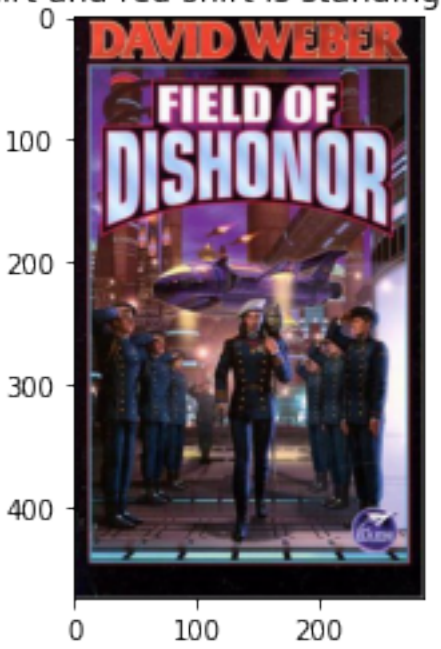
[146]:
```
# Performing image captioning for a single sample image from the dataset
img = 'fantasy20.jpg'
testingEnc[img] = encodeData('../input/sampleimage/' + img,featExtractor)
```

[147]:
```
feature_vector = testingEnc[img].reshape((1,2048))
x = plt.imread('../input/sampleimage/fantasy20.jpg')
plt.imshow(x)
plt.title(greedy_search(feature_vector,0))
plt.show()
```



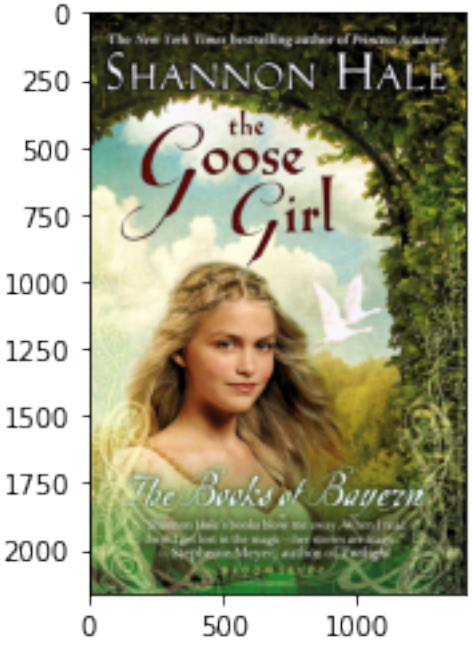
[148]:
```
# Performing image captioning for several sample images from the dataset
for myfile in os.listdir('../input/sampleimage/'):
    img = myfile
    encoded_data = encodeData('../input/sampleimage/' + img,featExtractor)
    feature_vector = encoded_data.reshape((1,2048))
    x = plt.imread('../input/sampleimage/'+img)
    plt.imshow(x)
    plt.title(greedy_search(feature_vector,0))
    plt.show()
```

man in red shirt and red shirt is standing in front of crowd

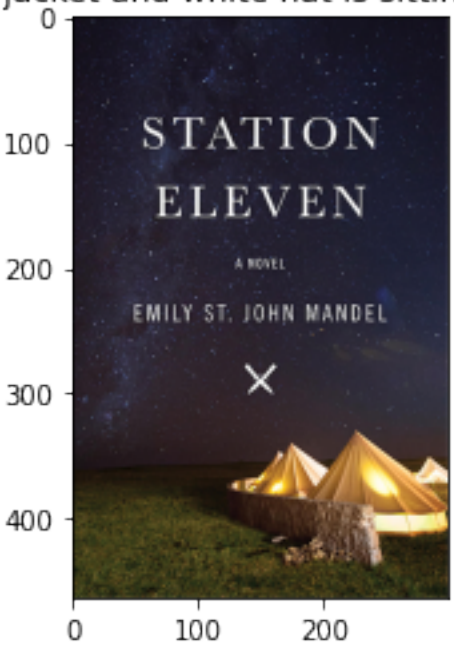


two women are sitting on the floor

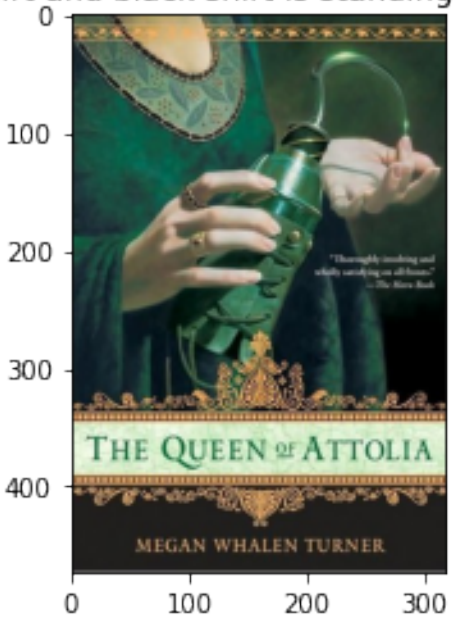
woman in red shirt and white shirt is sitting on the sidewalk



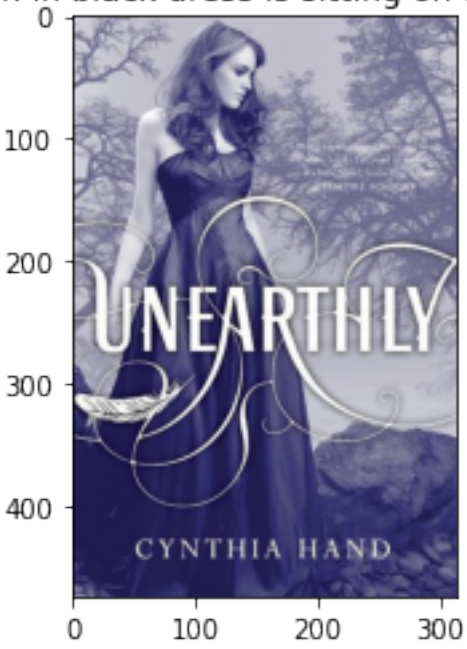
man in red jacket and white hat is sitting on the beach



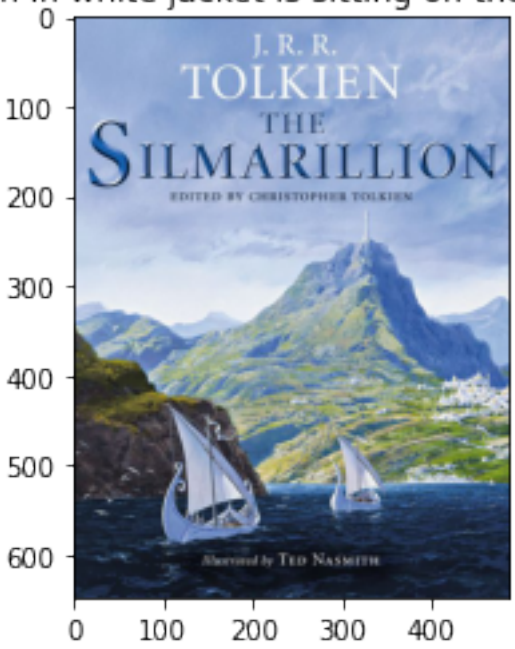
woman in black shirt and black shirt is standing in front of the camera



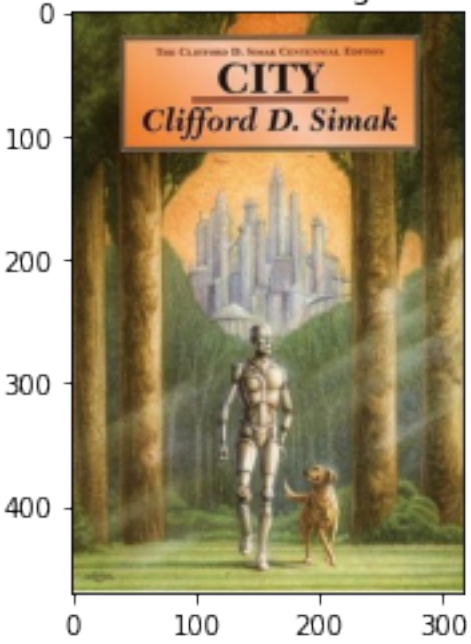
woman in black dress is sitting on the beach



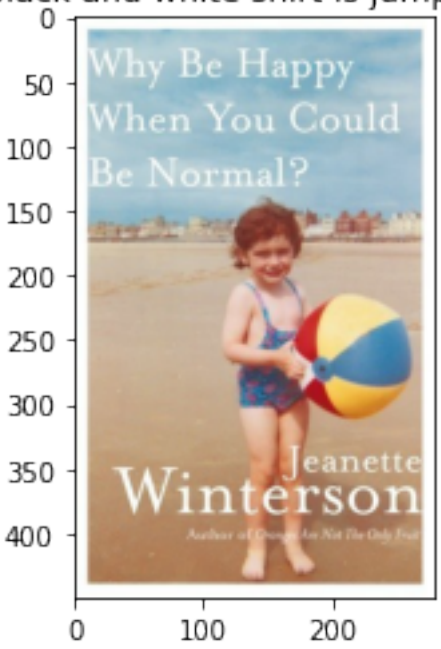
man in white jacket is sitting on the beach



man and woman are sitting on the ground

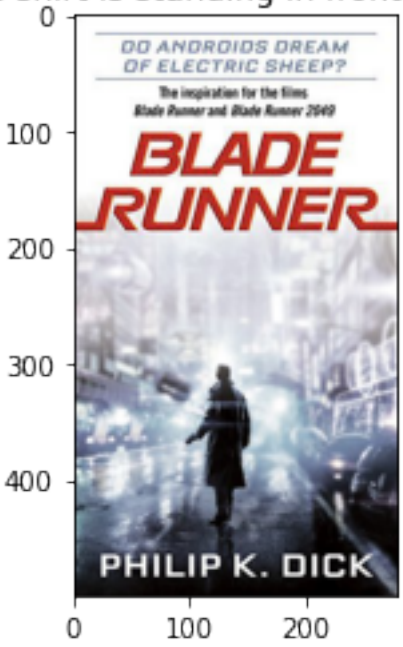


woman in black and white shirt is jumping into the air

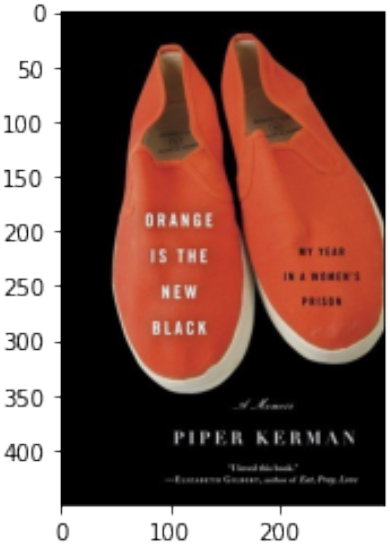


woman in black dress and white shirt is sitting on the street

man in red shirt is standing in front of the camera

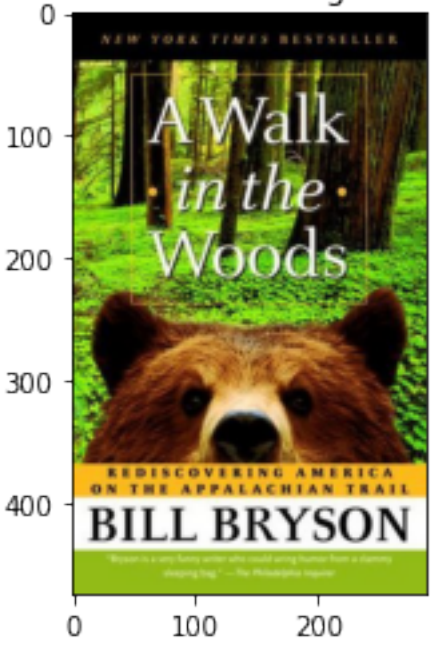


woman wearing red shirt and black shirt and white shirt is sitting on the ground

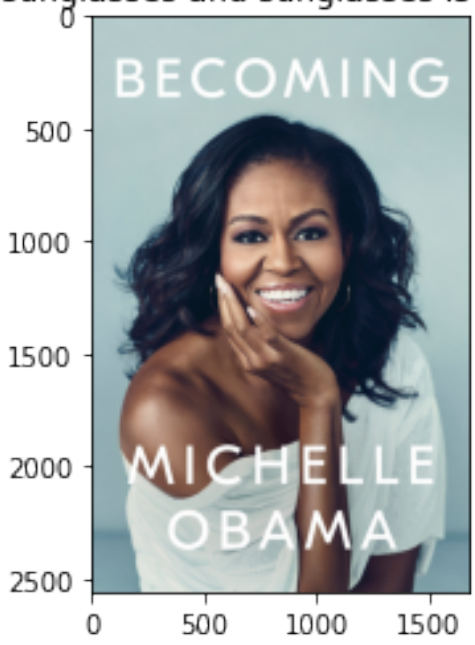


woman wearing red shirt and white shirt and white shirt is holding her head

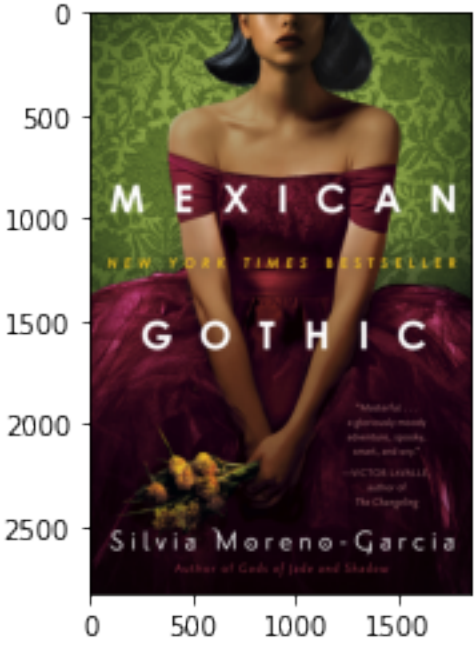
man in red shirt is sitting on the grass



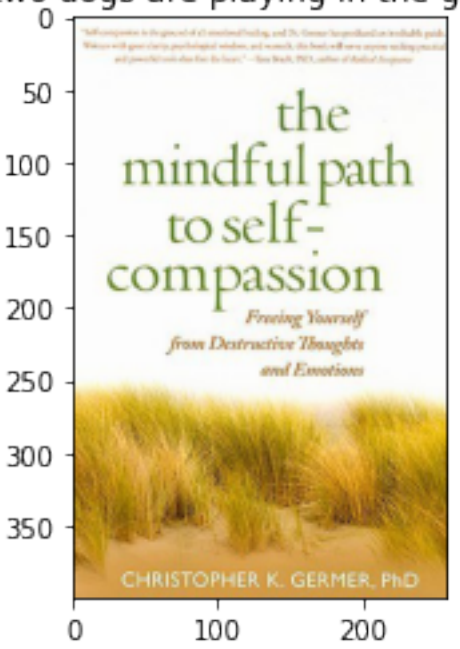
woman wearing sunglasses and sunglasses is sitting on the camera



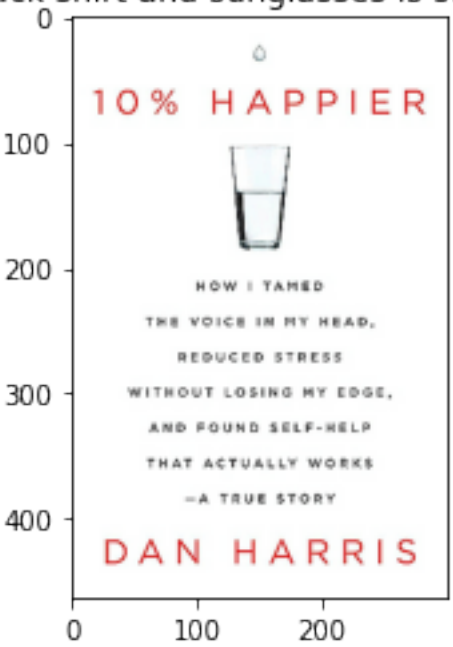
woman in black dress and black hat and woman in black dress



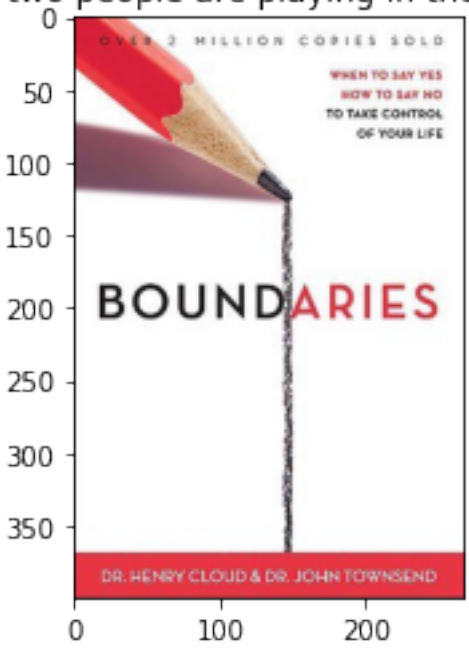
two dogs are playing in the grass



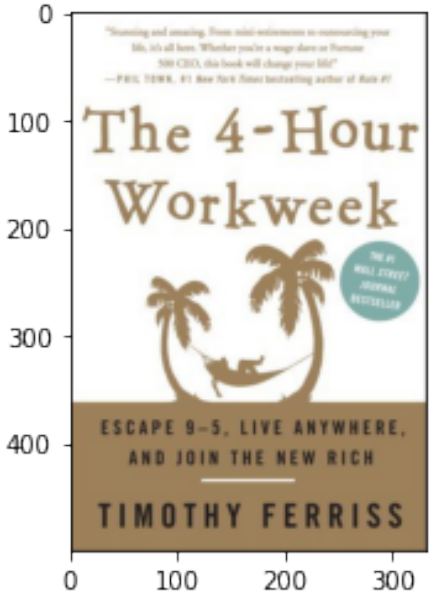
man wearing black shirt and sunglasses is sitting on the camera



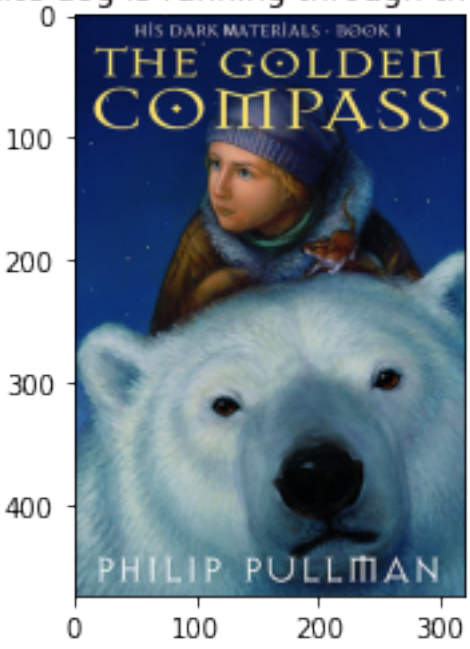
two people are playing in the air



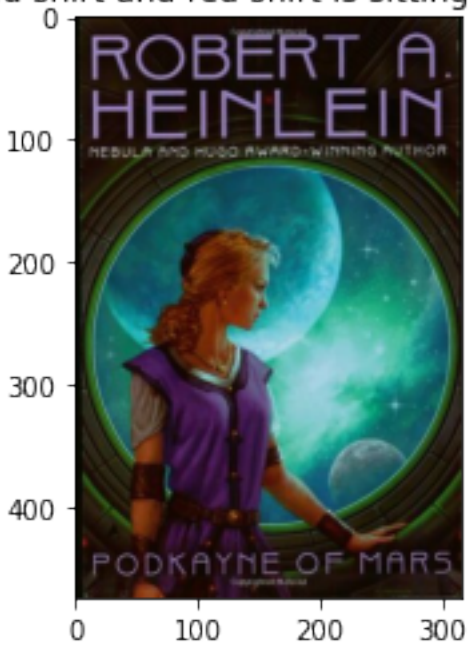
man in black shirt and black shirt and white shirt is sitting on the ground



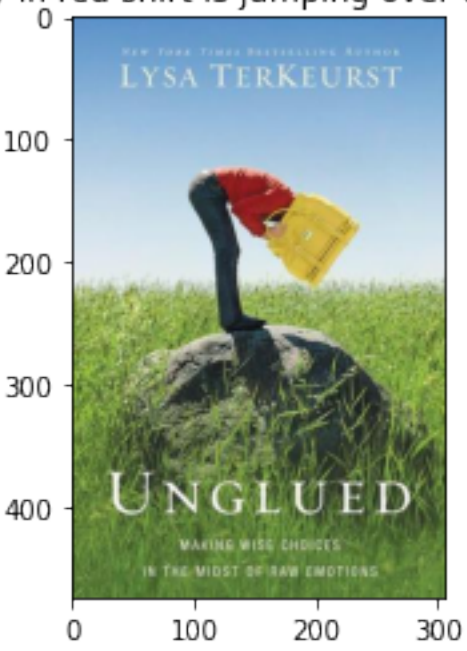
white dog is running through the snow



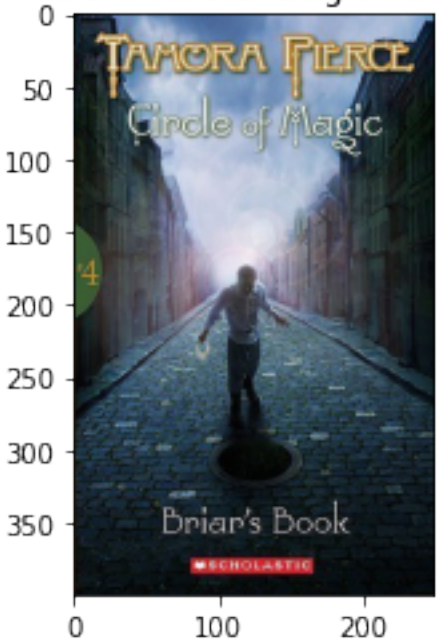
man in red shirt and red shirt is sitting on the street



boy in red shirt is jumping over the air



man in red shirt is standing on the sidewalk



woman wearing red shirt and sunglasses and white shirt and white shirt and white shirt and white shirt and white shirt and white shirt and white