Image Captioning using Flickr8k Dataset

GROUP 6 - Final Project
DATS 6203(10)
Machine Learning II

Instructor: Prof. Amir Jafari

Individual Final Report

By: Madhuri Yadav
(December 7, 2020)

Table of Contents

Introduction:

This project aims to achieve Image Captioning using Deep Learning Techniques, mainly Convolutional Neural Networks and Long Short-Term Memory. The dataset used for this project is the Flickr8k dataset, found easily on Kaggle. Flickr8k, and others like it, such as the COCO dataset, have created the benchmark for sentence-based image captioning.

The scope of the project incorporates different domains under Artificial Intelligence such as Computer Vision, Deep Learning, and Natural Language Processing. Applications and business value of this project include:

- Image Search in Search Engines
- Image Segregation and Classification
- Text-to-Speech to aid visually impaired individuals
- Automatic Image annotation in Facial Recognition, E-commerce, etc.

The project is divided into 5 main components:

1. Data Splitting – generating the train and test ids of the images
2. Text Preprocessing – preprocessing the captions before being fed into the LSTM model
3. Image Preprocessing – generating image features using CNN models
4. LSTM – Caption generator – generates captions
5. Model Evaluation – using various performance metrics

Dataset Description

The Flickr8k dataset has 2 main components – Images and their associated Captions. There are 8092 images, each having 5 captions in the captions.txt file. Hence, there are a total of 40,460 captions. The dataset can be found here.

Architecture:

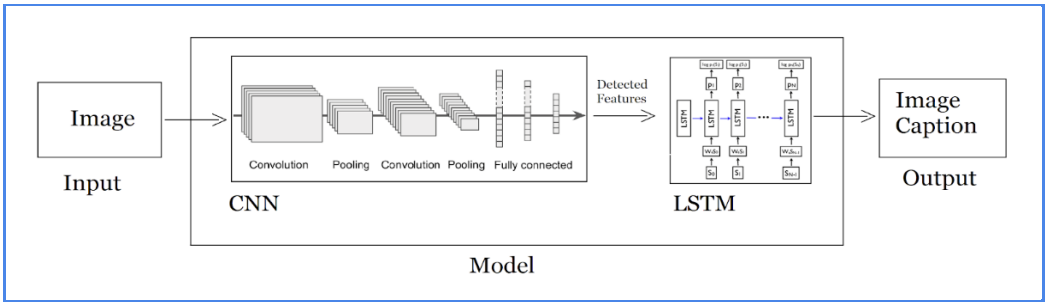

Figure 1: Showing the Architecture used for the project

We are using an Encoder-Decoder model as per our base paper to tackle this problem. Here our encoder model will combine both the encoded form of the image and the encoded form of the text caption and feed to the decoder. Our model will treat CNN as the 'image model' and the RNN/LSTM as the 'language model' to encode the text sequences of varying length. The vectors resulting from both the encodings are then merged and processed by a Dense layer to make a final prediction.

Description of Individual Work:

I have contributed to the Data splitting and the Image preprocessing (image encoder) part:

Data Split:

For Splitting the Dataset list of images is taken from the dataset folder and the list is divided into 80%-20% train and test respectively. The code generates two files train_ids.txt and test_ids.txt which has the list of image ids/names.

Image Preprocessing:

Before passing the images to the encoder images are preprocessed. The train image ids received from the previous step are read id by id, and each image is then resized to the input size used by CNN model (244X244 for VGG-16). The resized image is then converted to array and reshaped to pass it to the CNN model which return the encoded image feature to pass to the LSTM model.

Similar process is followed with the test id's and the encoded features are passed to the prediction model.

Image Model: (Features Generator)

Convoluted neural Network works here as an encoder model. As we know since CNN uses Kernels, the model could be used to extract features from the images and this step we denote as an image encoding. There are a lot of models that we can use like VGG-16, InceptionV3, ResNet, etc. Which are generally used for image captioning. We tried All the three models for our project to see which encoder gives the best results. Since we found VGG-16 gives the best results with our data set let us discuss the details of the model.

VGG-16: Our Best Feature Generator

VGG16 is a convolutional neural network model proposed by K. Simonyan and A. Zisserman. The model achieves 92.7% top-5 test accuracy in ImageNet, which is a dataset of over 14 million images belonging to 1000 classes.

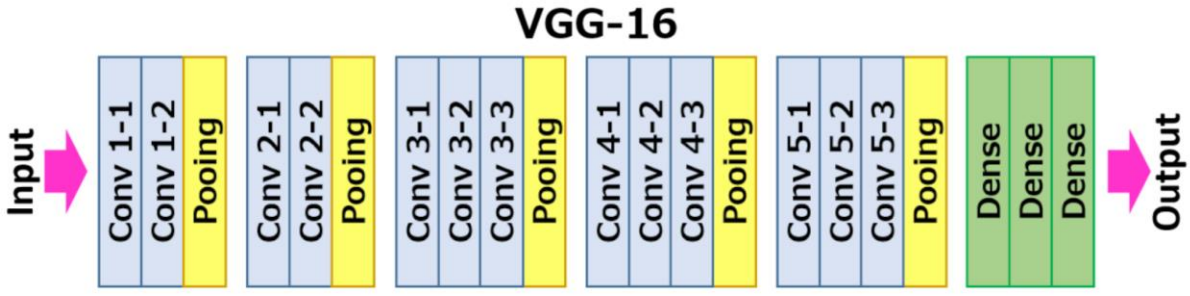

Figure 2: Layers of VGG-16

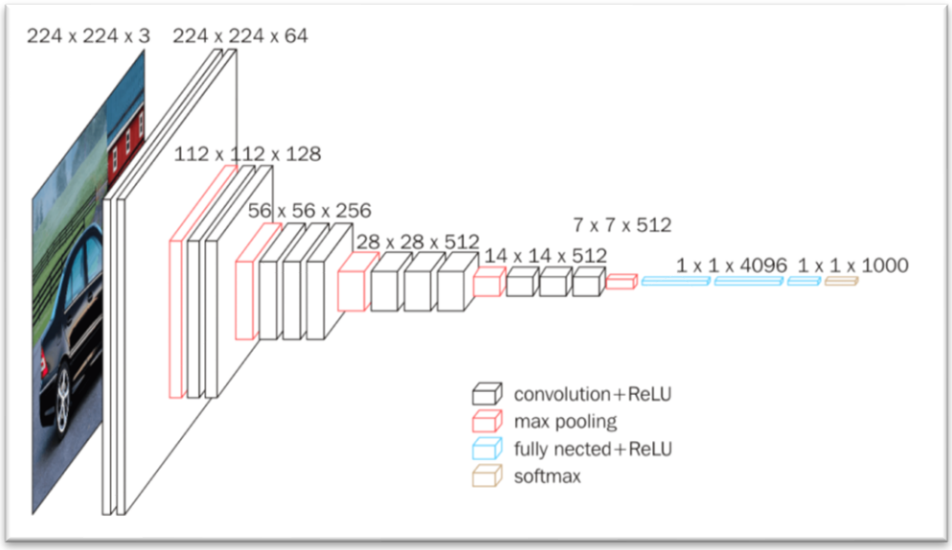

Figure 3: The architecture of the VGG-16.(With final 2 dense layers.o)

The Model is named as VGG-16 as it uses 16 hidden layers in its architecture. The input to cov1 layer is of fixed size 224 x 224 RGB image. The image is passed through a stack of convolutional layers, where the filters with very small receptive field: 3×3 & 1×1 are used. The network uses Max-pooling with stride of 2. It has Three Fully-Connected (FC) layers the first two have 4096 channels each. The third layer is the soft-max layer contains 1000 channels. The model has 138,357,544 parameters in total. The summary of the model is as follows.

```
Model: "vgg16"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_1 (InputLayer)         [(None, 224, 224, 3)]     0
_________________________________________________________________
block1_conv1 (Conv2D)        (None, 224, 224, 64)      1792
_________________________________________________________________
block1_conv2 (Conv2D)        (None, 224, 224, 64)      36928
_________________________________________________________________
block1_pool (MaxPooling2D)   (None, 112, 112, 64)      0
_________________________________________________________________
block2_conv1 (Conv2D)        (None, 112, 112, 128)     73856
_________________________________________________________________
block2_conv2 (Conv2D)        (None, 112, 112, 128)     147584
_________________________________________________________________
block2_pool (MaxPooling2D)   (None, 56, 56, 128)       0
_________________________________________________________________
block3_conv1 (Conv2D)        (None, 56, 56, 256)       295168
_________________________________________________________________
block3_conv2 (Conv2D)        (None, 56, 56, 256)       590080
_________________________________________________________________
block3_conv3 (Conv2D)        (None, 56, 56, 256)       590080
_________________________________________________________________
block3_pool (MaxPooling2D)   (None, 28, 28, 256)       0
_________________________________________________________________
```

```
block4_conv1 (Conv2D)          (None, 28, 28, 512)         1180160
_________________________________________________________________
block4_conv2 (Conv2D)          (None, 28, 28, 512)         2359808
_________________________________________________________________
block4_conv3 (Conv2D)          (None, 28, 28, 512)         2359808
_________________________________________________________________
block4_pool (MaxPooling2D)     (None, 14, 14, 512)         0
_________________________________________________________________
block5_conv1 (Conv2D)          (None, 14, 14, 512)         2359808
_________________________________________________________________
block5_conv2 (Conv2D)          (None, 14, 14, 512)         2359808
_________________________________________________________________
block5_conv3 (Conv2D)          (None, 14, 14, 512)         2359808
_________________________________________________________________
block5_pool (MaxPooling2D)     (None, 7, 7, 512)           0
_________________________________________________________________
flatten (Flatten)              (None, 25088)               0
_________________________________________________________________
fc1 (Dense)                    (None, 4096)                102764544
_________________________________________________________________
fc2 (Dense)                    (None, 4096)                16781312
_________________________________________________________________
predictions (Dense)            (None, 1000)                4097000
=================================================================
Total params: 138,357,544
Trainable params: 138,357,544
Non-trainable params: 0
_________________________________________________________________
```

As the purpose of using CNN here is not for image classification and is for image extraction, we eliminate last 2 dense layers to get the flattened features which could be passed to our RNN/LSTM (Decoder) model.

Other than VGG-16 initially we tried to use our own networks, which was not a good encoder, hence based on our research on image captioning we also tried InceptionV3 (Size: 92MB, Top-5 Acc: 0.937, Parameters: 23,851,784), ResNet50V2 (Size: 98MB, Top-5 Acc: 0.93, Parameters: 25,613,800). These models are also well-known feature extractors for image captioning purposes.

InceptionV3:

The model looks like shown in the figure 4.

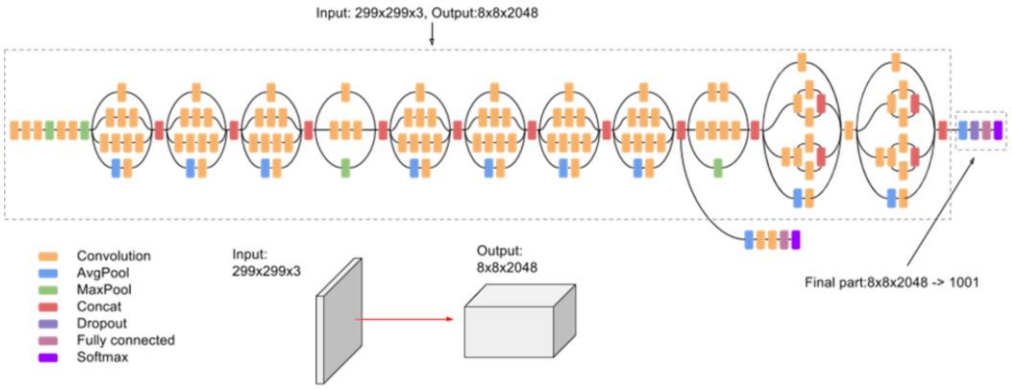

Figure 4: Inception V3 architecture

As we can see from the figure 3. The model uses several Conv layers mainly uses avg pooling and has 23,851,784 parameters. The model takes input of 299X299 image size and after we eliminate the final dense layer (Softmax) the model has output size of 2048. The model summary in the Appendix A.

ResNet50V2:

The model looks like shown in the figure 5.

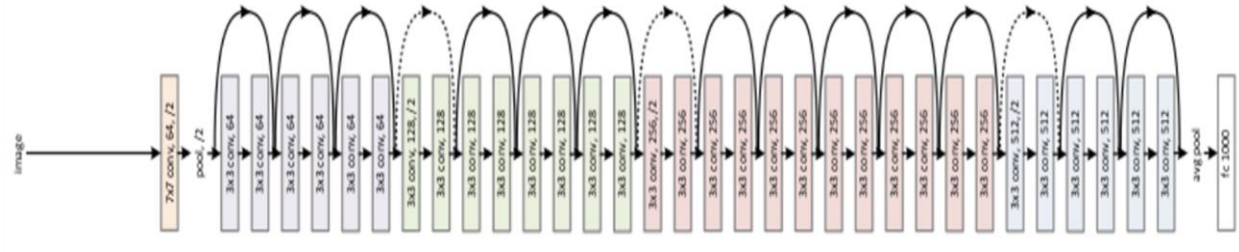

Figure 5: ResNet50V2 Architecture

As we can see from the figure 3. The ResNet-50 model consists of 5 stages each with a convolution and Identity block. Each convolution block has 3 convolution layers and each identity block also has 3 convolution layers and has 25,613,800 parameters. The model Solves the vanishing gradient problem as it uses multiple layers. The model takes input of fixed size 224 x 224 RGB image and after we eliminate the final dense layer (Softmax) the model has output size of 2048. The model summary in the Appendix B.

Feature extraction:

extract_features method implemented generates a pickle file. The pickle file contains image features dictionary with image name as key and features generated by the CNN model (size 4096 each) as its value. The Pickle file is then passed to the LSTM model and extracted as dictionary back.

Results:

We used the BLEU score metric to evaluate performance of our model. A perfect match results in a score of 1.0, whereas a perfect mismatch results in a score of 0.0. The approach works by counting matching n-grams in the candidate translation to n-grams in the reference text. Below are the cumulative BLEU scores on test set for feature extraction with VGG-16 model.

```
BLEU-1: 0.412581
BLEU-2: 0.230329
BLEU-3: 0.154373
BLEU-4: 0.070913
```

Summary and Conclusions:

With this we conclude that for our dataset VGG-16 works the best as feature extraction. InceptionV3 and ResNet50V2 Mostly tend to overfit the data and as a result we get similar output for majority images. The results are observed to be much better with the first model which gives a BLUE-1 of 0.41.

To implement this code for the final model I added around 50% of the code. For the image feature extraction and data splitting.

References.

1. https://www.analyticsvidhya.com/blog/2020/11/create-your-own-image-caption-generator-using-keras/
2. https://neurohive.io/en/popular-networks/vgg16/
3. https://towardsdatascience.com/review-resnet-winner-of-ilsvrc-2015-image-classification-localization-detection-e39402bfa5d8
4. https://cloud.google.com/tpu/docs/inception-v3-advanced

Appendix A: Inception v3 model Summary

Model: "inception_v3"

Layer (type)	Output Shape	Param #	Connected to
input_1 (InputLayer)	[(None, 299, 299, 3)	0	
conv2d (Conv2D)	(None, 149, 149, 32)	864	input_1[0][0]
batch_normalization (BatchNorma	(None, 149, 149, 32)	96	conv2d[0][0]
activation (Activation)	(None, 149, 149, 32)	0	batch_normalization[0][0]
conv2d_1 (Conv2D)	(None, 147, 147, 32)	9216	activation[0][0]
batch_normalization_1 (BatchNor	(None, 147, 147, 32)	96	conv2d_1[0][0]
activation_1 (Activation)	(None, 147, 147, 32)	0	batch_normalization_1[0][0]
conv2d_2 (Conv2D)	(None, 147, 147, 64)	18432	activation_1[0][0]
batch_normalization_2 (BatchNor	(None, 147, 147, 64)	192	conv2d_2[0][0]
activation_2 (Activation)	(None, 147, 147, 64)	0	batch_normalization_2[0][0]
max_pooling2d (MaxPooling2D)	(None, 73, 73, 64)	0	activation_2[0][0]
conv2d_3 (Conv2D)	(None, 73, 73, 80)	5120	max_pooling2d[0][0]
batch_normalization_3 (BatchNor	(None, 73, 73, 80)	240	conv2d_3[0][0]
activation_3 (Activation)	(None, 73, 73, 80)	0	batch_normalization_3[0][0]
conv2d_4 (Conv2D)	(None, 71, 71, 192)	138240	activation_3[0][0]
batch_normalization_4 (BatchNor	(None, 71, 71, 192)	576	conv2d_4[0][0]
activation_4 (Activation)	(None, 71, 71, 192)	0	batch_normalization_4[0][0]
max_pooling2d_1 (MaxPooling2D)	(None, 35, 35, 192)	0	activation_4[0][0]
conv2d_8 (Conv2D)	(None, 35, 35, 64)	12288	max_pooling2d_1[0][0]
batch_normalization_8 (BatchNor	(None, 35, 35, 64)	192	conv2d_8[0][0]
activation_8 (Activation)	(None, 35, 35, 64)	0	batch_normalization_8[0][0]
conv2d_6 (Conv2D)	(None, 35, 35, 48)	9216	max_pooling2d_1[0][0]
conv2d_9 (Conv2D)	(None, 35, 35, 96)	55296	activation_8[0][0]

```
batch_normalization_6 (BatchNor (None, 35, 35, 48)   144         conv2d_6[0][0]
batch_normalization_9 (BatchNor (None, 35, 35, 96)   288         conv2d_9[0][0]
activation_6 (Activation)       (None, 35, 35, 48)   0           batch_normalization_6[0][0]
activation_9 (Activation)       (None, 35, 35, 96)   0           batch_normalization_9[0][0]
average_pooling2d (AveragePooli (None, 35, 35, 192)  0           max_pooling2d_1[0][0]
conv2d_5 (Conv2D)               (None, 35, 35, 64)   12288       max_pooling2d_1[0][0]
conv2d_7 (Conv2D)               (None, 35, 35, 64)   76800       activation_6[0][0]
conv2d_10 (Conv2D)              (None, 35, 35, 96)   82944       activation_9[0][0]
conv2d_11 (Conv2D)              (None, 35, 35, 32)   6144        average_pooling2d[0][0]
batch_normalization_5 (BatchNor (None, 35, 35, 64)   192         conv2d_5[0][0]
batch_normalization_7 (BatchNor (None, 35, 35, 64)   192         conv2d_7[0][0]
batch_normalization_10 (BatchNo (None, 35, 35, 96)   288         conv2d_10[0][0]
batch_normalization_11 (BatchNo (None, 35, 35, 32)   96          conv2d_11[0][0]
activation_5 (Activation)       (None, 35, 35, 64)   0           batch_normalization_5[0][0]
activation_7 (Activation)       (None, 35, 35, 64)   0           batch_normalization_7[0][0]
activation_10 (Activation)      (None, 35, 35, 96)   0           batch_normalization_10[0][0]
activation_11 (Activation)      (None, 35, 35, 32)   0           batch_normalization_11[0][0]
mixed0 (Concatenate)            (None, 35, 35, 256)  0           activation_5[0][0]
                                                                 activation_7[0][0]
                                                                 activation_10[0][0]
                                                                 activation_11[0][0]
conv2d_15 (Conv2D)              (None, 35, 35, 64)   16384       mixed0[0][0]
batch_normalization_15 (BatchNo (None, 35, 35, 64)   192         conv2d_15[0][0]
activation_15 (Activation)      (None, 35, 35, 64)   0           batch_normalization_15[0][0]
conv2d_13 (Conv2D)              (None, 35, 35, 48)   12288       mixed0[0][0]
conv2d_16 (Conv2D)              (None, 35, 35, 96)   55296       activation_15[0][0]
batch_normalization_13 (BatchNo (None, 35, 35, 48)   144         conv2d_13[0][0]
batch_normalization_16 (BatchNo (None, 35, 35, 96)   288         conv2d_16[0][0]
activation_13 (Activation)      (None, 35, 35, 48)   0           batch_normalization_13[0][0]
activation_16 (Activation)      (None, 35, 35, 96)   0           batch_normalization_16[0][0]
average_pooling2d_1 (AveragePoo (None, 35, 35, 256)  0           mixed0[0][0]
conv2d_12 (Conv2D)              (None, 35, 35, 64)   16384       mixed0[0][0]
conv2d_14 (Conv2D)              (None, 35, 35, 64)   76800       activation_13[0][0]
conv2d_17 (Conv2D)              (None, 35, 35, 96)   82944       activation_16[0][0]
conv2d_18 (Conv2D)              (None, 35, 35, 64)   16384       average_pooling2d_1[0][0]
batch_normalization_12 (BatchNo (None, 35, 35, 64)   192         conv2d_12[0][0]
batch_normalization_14 (BatchNo (None, 35, 35, 64)   192         conv2d_14[0][0]
batch_normalization_17 (BatchNo (None, 35, 35, 96)   288         conv2d_17[0][0]
batch_normalization_18 (BatchNo (None, 35, 35, 64)   192         conv2d_18[0][0]
activation_12 (Activation)      (None, 35, 35, 64)   0           batch_normalization_12[0][0]
activation_14 (Activation)      (None, 35, 35, 64)   0           batch_normalization_14[0][0]
activation_17 (Activation)      (None, 35, 35, 96)   0           batch_normalization_17[0][0]
activation_18 (Activation)      (None, 35, 35, 64)   0           batch_normalization_18[0][0]
mixed1 (Concatenate)            (None, 35, 35, 288)  0           activation_12[0][0]
                                                                 activation_14[0][0]
                                                                 activation_17[0][0]
                                                                 activation_18[0][0]
```

```
conv2d_22 (Conv2D)              (None, 35, 35, 64)   18432       mixed1[0][0]
batch_normalization_22 (BatchNo (None, 35, 35, 64)   192         conv2d_22[0][0]
activation_22 (Activation)      (None, 35, 35, 64)   0           batch_normalization_22[0][0]
conv2d_20 (Conv2D)              (None, 35, 35, 48)   13824       mixed1[0][0]
conv2d_23 (Conv2D)              (None, 35, 35, 96)   55296       activation_22[0][0]
batch_normalization_20 (BatchNo (None, 35, 35, 48)   144         conv2d_20[0][0]
batch_normalization_23 (BatchNo (None, 35, 35, 96)   288         conv2d_23[0][0]
activation_20 (Activation)      (None, 35, 35, 48)   0           batch_normalization_20[0][0]
activation_23 (Activation)      (None, 35, 35, 96)   0           batch_normalization_23[0][0]
average_pooling2d_2 (AveragePoo (None, 35, 35, 288)  0           mixed1[0][0]
conv2d_19 (Conv2D)              (None, 35, 35, 64)   18432       mixed1[0][0]
conv2d_21 (Conv2D)              (None, 35, 35, 64)   76800       activation_20[0][0]
conv2d_24 (Conv2D)              (None, 35, 35, 96)   82944       activation_23[0][0]
conv2d_25 (Conv2D)              (None, 35, 35, 64)   18432       average_pooling2d_2[0][0]
batch_normalization_19 (BatchNo (None, 35, 35, 64)   192         conv2d_19[0][0]
batch_normalization_21 (BatchNo (None, 35, 35, 64)   192         conv2d_21[0][0]
batch_normalization_24 (BatchNo (None, 35, 35, 96)   288         conv2d_24[0][0]
batch_normalization_25 (BatchNo (None, 35, 35, 64)   192         conv2d_25[0][0]
activation_19 (Activation)      (None, 35, 35, 64)   0           batch_normalization_19[0][0]
activation_21 (Activation)      (None, 35, 35, 64)   0           batch_normalization_21[0][0]
activation_24 (Activation)      (None, 35, 35, 96)   0           batch_normalization_24[0][0]
activation_25 (Activation)      (None, 35, 35, 64)   0           batch_normalization_25[0][0]
mixed2 (Concatenate)            (None, 35, 35, 288)  0           activation_19[0][0]
                                                                 activation_21[0][0]
                                                                 activation_24[0][0]
                                                                 activation_25[0][0]
conv2d_27 (Conv2D)              (None, 35, 35, 64)   18432       mixed2[0][0]
batch_normalization_27 (BatchNo (None, 35, 35, 64)   192         conv2d_27[0][0]
activation_27 (Activation)      (None, 35, 35, 64)   0           batch_normalization_27[0][0]
conv2d_28 (Conv2D)              (None, 35, 35, 96)   55296       activation_27[0][0]
batch_normalization_28 (BatchNo (None, 35, 35, 96)   288         conv2d_28[0][0]
activation_28 (Activation)      (None, 35, 35, 96)   0           batch_normalization_28[0][0]
conv2d_26 (Conv2D)              (None, 17, 17, 384)  995328      mixed2[0][0]
conv2d_29 (Conv2D)              (None, 17, 17, 96)   82944       activation_28[0][0]
batch_normalization_26 (BatchNo (None, 17, 17, 384)  1152        conv2d_26[0][0]
batch_normalization_29 (BatchNo (None, 17, 17, 96)   288         conv2d_29[0][0]
activation_26 (Activation)      (None, 17, 17, 384)  0           batch_normalization_26[0][0]
activation_29 (Activation)      (None, 17, 17, 96)   0           batch_normalization_29[0][0]
max_pooling2d_2 (MaxPooling2D)  (None, 17, 17, 288)  0           mixed2[0][0]
mixed3 (Concatenate)            (None, 17, 17, 768)  0           activation_26[0][0]
                                                                 activation_29[0][0]
                                                                 max_pooling2d_2[0][0]
conv2d_34 (Conv2D)              (None, 17, 17, 128)  98304       mixed3[0][0]
batch_normalization_34 (BatchNo (None, 17, 17, 128)  384         conv2d_34[0][0]
activation_34 (Activation)      (None, 17, 17, 128)  0           batch_normalization_34[0][0]
conv2d_35 (Conv2D)              (None, 17, 17, 128)  114688      activation_34[0][0]
```

```
batch_normalization_35 (BatchNo (None, 17, 17, 128)  384         conv2d_35[0][0]
__________________________________________________________________________________________________
activation_35 (Activation)      (None, 17, 17, 128)  0           batch_normalization_35[0][0]
__________________________________________________________________________________________________
conv2d_31 (Conv2D)              (None, 17, 17, 128)  98304       mixed3[0][0]
__________________________________________________________________________________________________
conv2d_36 (Conv2D)              (None, 17, 17, 128)  114688      activation_35[0][0]
__________________________________________________________________________________________________
batch_normalization_31 (BatchNo (None, 17, 17, 128)  384         conv2d_31[0][0]
__________________________________________________________________________________________________
batch_normalization_36 (BatchNo (None, 17, 17, 128)  384         conv2d_36[0][0]
__________________________________________________________________________________________________
activation_31 (Activation)      (None, 17, 17, 128)  0           batch_normalization_31[0][0]
__________________________________________________________________________________________________
activation_36 (Activation)      (None, 17, 17, 128)  0           batch_normalization_36[0][0]
__________________________________________________________________________________________________
conv2d_32 (Conv2D)              (None, 17, 17, 128)  114688      activation_31[0][0]
__________________________________________________________________________________________________
conv2d_37 (Conv2D)              (None, 17, 17, 128)  114688      activation_36[0][0]
__________________________________________________________________________________________________
batch_normalization_32 (BatchNo (None, 17, 17, 128)  384         conv2d_32[0][0]
__________________________________________________________________________________________________
batch_normalization_37 (BatchNo (None, 17, 17, 128)  384         conv2d_37[0][0]
__________________________________________________________________________________________________
activation_32 (Activation)      (None, 17, 17, 128)  0           batch_normalization_32[0][0]
__________________________________________________________________________________________________
activation_37 (Activation)      (None, 17, 17, 128)  0           batch_normalization_37[0][0]
__________________________________________________________________________________________________
average_pooling2d_3 (AveragePoo (None, 17, 17, 768)  0           mixed3[0][0]
__________________________________________________________________________________________________
conv2d_30 (Conv2D)              (None, 17, 17, 192)  147456      mixed3[0][0]
__________________________________________________________________________________________________
conv2d_33 (Conv2D)              (None, 17, 17, 192)  172032      activation_32[0][0]
__________________________________________________________________________________________________
conv2d_38 (Conv2D)              (None, 17, 17, 192)  172032      activation_37[0][0]
__________________________________________________________________________________________________
conv2d_39 (Conv2D)              (None, 17, 17, 192)  147456      average_pooling2d_3[0][0]
__________________________________________________________________________________________________
batch_normalization_30 (BatchNo (None, 17, 17, 192)  576         conv2d_30[0][0]
__________________________________________________________________________________________________
batch_normalization_33 (BatchNo (None, 17, 17, 192)  576         conv2d_33[0][0]
__________________________________________________________________________________________________
batch_normalization_38 (BatchNo (None, 17, 17, 192)  576         conv2d_38[0][0]
__________________________________________________________________________________________________
batch_normalization_39 (BatchNo (None, 17, 17, 192)  576         conv2d_39[0][0]
__________________________________________________________________________________________________
activation_30 (Activation)      (None, 17, 17, 192)  0           batch_normalization_30[0][0]
__________________________________________________________________________________________________
activation_33 (Activation)      (None, 17, 17, 192)  0           batch_normalization_33[0][0]
__________________________________________________________________________________________________
activation_38 (Activation)      (None, 17, 17, 192)  0           batch_normalization_38[0][0]
__________________________________________________________________________________________________
activation_39 (Activation)      (None, 17, 17, 192)  0           batch_normalization_39[0][0]
__________________________________________________________________________________________________
mixed4 (Concatenate)            (None, 17, 17, 768)  0           activation_30[0][0]
                                                                 activation_33[0][0]
                                                                 activation_38[0][0]
                                                                 activation_39[0][0]
__________________________________________________________________________________________________
conv2d_44 (Conv2D)              (None, 17, 17, 160)  122880      mixed4[0][0]
__________________________________________________________________________________________________
batch_normalization_44 (BatchNo (None, 17, 17, 160)  480         conv2d_44[0][0]
__________________________________________________________________________________________________
activation_44 (Activation)      (None, 17, 17, 160)  0           batch_normalization_44[0][0]
__________________________________________________________________________________________________
conv2d_45 (Conv2D)              (None, 17, 17, 160)  179200      activation_44[0][0]
__________________________________________________________________________________________________
batch_normalization_45 (BatchNo (None, 17, 17, 160)  480         conv2d_45[0][0]
__________________________________________________________________________________________________
activation_45 (Activation)      (None, 17, 17, 160)  0           batch_normalization_45[0][0]
__________________________________________________________________________________________________
conv2d_41 (Conv2D)              (None, 17, 17, 160)  122880      mixed4[0][0]
__________________________________________________________________________________________________
conv2d_46 (Conv2D)              (None, 17, 17, 160)  179200      activation_45[0][0]
__________________________________________________________________________________________________
batch_normalization_41 (BatchNo (None, 17, 17, 160)  480         conv2d_41[0][0]
__________________________________________________________________________________________________
batch_normalization_46 (BatchNo (None, 17, 17, 160)  480         conv2d_46[0][0]
__________________________________________________________________________________________________
activation_41 (Activation)      (None, 17, 17, 160)  0           batch_normalization_41[0][0]
__________________________________________________________________________________________________
activation_46 (Activation)      (None, 17, 17, 160)  0           batch_normalization_46[0][0]
__________________________________________________________________________________________________
conv2d_42 (Conv2D)              (None, 17, 17, 160)  179200      activation_41[0][0]
__________________________________________________________________________________________________
conv2d_47 (Conv2D)              (None, 17, 17, 160)  179200      activation_46[0][0]
__________________________________________________________________________________________________
batch_normalization_42 (BatchNo (None, 17, 17, 160)  480         conv2d_42[0][0]
```

```
batch_normalization_47 (BatchNo (None, 17, 17, 160)  480         conv2d_47[0][0]
activation_42 (Activation)      (None, 17, 17, 160)  0           batch_normalization_42[0][0]
activation_47 (Activation)      (None, 17, 17, 160)  0           batch_normalization_47[0][0]
average_pooling2d_4 (AveragePoo (None, 17, 17, 768)  0           mixed4[0][0]
conv2d_40 (Conv2D)              (None, 17, 17, 192)  147456      mixed4[0][0]
conv2d_43 (Conv2D)              (None, 17, 17, 192)  215040      activation_42[0][0]
conv2d_48 (Conv2D)              (None, 17, 17, 192)  215040      activation_47[0][0]
conv2d_49 (Conv2D)              (None, 17, 17, 192)  147456      average_pooling2d_4[0][0]
batch_normalization_40 (BatchNo (None, 17, 17, 192)  576         conv2d_40[0][0]
batch_normalization_43 (BatchNo (None, 17, 17, 192)  576         conv2d_43[0][0]
batch_normalization_48 (BatchNo (None, 17, 17, 192)  576         conv2d_48[0][0]
batch_normalization_49 (BatchNo (None, 17, 17, 192)  576         conv2d_49[0][0]
activation_40 (Activation)      (None, 17, 17, 192)  0           batch_normalization_40[0][0]
activation_43 (Activation)      (None, 17, 17, 192)  0           batch_normalization_43[0][0]
activation_48 (Activation)      (None, 17, 17, 192)  0           batch_normalization_48[0][0]
activation_49 (Activation)      (None, 17, 17, 192)  0           batch_normalization_49[0][0]
mixed5 (Concatenate)            (None, 17, 17, 768)  0           activation_40[0][0]
                                                                 activation_43[0][0]
                                                                 activation_48[0][0]
                                                                 activation_49[0][0]
conv2d_54 (Conv2D)              (None, 17, 17, 160)  122880      mixed5[0][0]
batch_normalization_54 (BatchNo (None, 17, 17, 160)  480         conv2d_54[0][0]
activation_54 (Activation)      (None, 17, 17, 160)  0           batch_normalization_54[0][0]
conv2d_55 (Conv2D)              (None, 17, 17, 160)  179200      activation_54[0][0]
batch_normalization_55 (BatchNo (None, 17, 17, 160)  480         conv2d_55[0][0]
activation_55 (Activation)      (None, 17, 17, 160)  0           batch_normalization_55[0][0]
conv2d_51 (Conv2D)              (None, 17, 17, 160)  122880      mixed5[0][0]
conv2d_56 (Conv2D)              (None, 17, 17, 160)  179200      activation_55[0][0]
batch_normalization_51 (BatchNo (None, 17, 17, 160)  480         conv2d_51[0][0]
batch_normalization_56 (BatchNo (None, 17, 17, 160)  480         conv2d_56[0][0]
activation_51 (Activation)      (None, 17, 17, 160)  0           batch_normalization_51[0][0]
activation_56 (Activation)      (None, 17, 17, 160)  0           batch_normalization_56[0][0]
conv2d_52 (Conv2D)              (None, 17, 17, 160)  179200      activation_51[0][0]
conv2d_57 (Conv2D)              (None, 17, 17, 160)  179200      activation_56[0][0]
batch_normalization_52 (BatchNo (None, 17, 17, 160)  480         conv2d_52[0][0]
batch_normalization_57 (BatchNo (None, 17, 17, 160)  480         conv2d_57[0][0]
activation_52 (Activation)      (None, 17, 17, 160)  0           batch_normalization_52[0][0]
activation_57 (Activation)      (None, 17, 17, 160)  0           batch_normalization_57[0][0]
average_pooling2d_5 (AveragePoo (None, 17, 17, 768)  0           mixed5[0][0]
conv2d_50 (Conv2D)              (None, 17, 17, 192)  147456      mixed5[0][0]
conv2d_53 (Conv2D)              (None, 17, 17, 192)  215040      activation_52[0][0]
conv2d_58 (Conv2D)              (None, 17, 17, 192)  215040      activation_57[0][0]
conv2d_59 (Conv2D)              (None, 17, 17, 192)  147456      average_pooling2d_5[0][0]
batch_normalization_50 (BatchNo (None, 17, 17, 192)  576         conv2d_50[0][0]
batch_normalization_53 (BatchNo (None, 17, 17, 192)  576         conv2d_53[0][0]
```

```
batch_normalization_58 (BatchNo (None, 17, 17, 192)  576         conv2d_58[0][0]
__________________________________________________________________________________________________
batch_normalization_59 (BatchNo (None, 17, 17, 192)  576         conv2d_59[0][0]
__________________________________________________________________________________________________
activation_50 (Activation)      (None, 17, 17, 192)  0           batch_normalization_50[0][0]
__________________________________________________________________________________________________
activation_53 (Activation)      (None, 17, 17, 192)  0           batch_normalization_53[0][0]
__________________________________________________________________________________________________
activation_58 (Activation)      (None, 17, 17, 192)  0           batch_normalization_58[0][0]
__________________________________________________________________________________________________
activation_59 (Activation)      (None, 17, 17, 192)  0           batch_normalization_59[0][0]
__________________________________________________________________________________________________
mixed6 (Concatenate)            (None, 17, 17, 768)  0           activation_50[0][0]
                                                                 activation_53[0][0]
                                                                 activation_58[0][0]
                                                                 activation_59[0][0]
__________________________________________________________________________________________________
conv2d_64 (Conv2D)              (None, 17, 17, 192)  147456      mixed6[0][0]
__________________________________________________________________________________________________
batch_normalization_64 (BatchNo (None, 17, 17, 192)  576         conv2d_64[0][0]
__________________________________________________________________________________________________
activation_64 (Activation)      (None, 17, 17, 192)  0           batch_normalization_64[0][0]
__________________________________________________________________________________________________
conv2d_65 (Conv2D)              (None, 17, 17, 192)  258048      activation_64[0][0]
__________________________________________________________________________________________________
batch_normalization_65 (BatchNo (None, 17, 17, 192)  576         conv2d_65[0][0]
__________________________________________________________________________________________________
activation_65 (Activation)      (None, 17, 17, 192)  0           batch_normalization_65[0][0]
__________________________________________________________________________________________________
conv2d_61 (Conv2D)              (None, 17, 17, 192)  147456      mixed6[0][0]
__________________________________________________________________________________________________
conv2d_66 (Conv2D)              (None, 17, 17, 192)  258048      activation_65[0][0]
__________________________________________________________________________________________________
batch_normalization_61 (BatchNo (None, 17, 17, 192)  576         conv2d_61[0][0]
__________________________________________________________________________________________________
batch_normalization_66 (BatchNo (None, 17, 17, 192)  576         conv2d_66[0][0]
__________________________________________________________________________________________________
activation_61 (Activation)      (None, 17, 17, 192)  0           batch_normalization_61[0][0]
__________________________________________________________________________________________________
activation_66 (Activation)      (None, 17, 17, 192)  0           batch_normalization_66[0][0]
__________________________________________________________________________________________________
conv2d_62 (Conv2D)              (None, 17, 17, 192)  258048      activation_61[0][0]
__________________________________________________________________________________________________
conv2d_67 (Conv2D)              (None, 17, 17, 192)  258048      activation_66[0][0]
__________________________________________________________________________________________________
batch_normalization_62 (BatchNo (None, 17, 17, 192)  576         conv2d_62[0][0]
__________________________________________________________________________________________________
batch_normalization_67 (BatchNo (None, 17, 17, 192)  576         conv2d_67[0][0]
__________________________________________________________________________________________________
activation_62 (Activation)      (None, 17, 17, 192)  0           batch_normalization_62[0][0]
__________________________________________________________________________________________________
activation_67 (Activation)      (None, 17, 17, 192)  0           batch_normalization_67[0][0]
__________________________________________________________________________________________________
average_pooling2d_6 (AveragePoo (None, 17, 17, 768)  0           mixed6[0][0]
__________________________________________________________________________________________________
conv2d_60 (Conv2D)              (None, 17, 17, 192)  147456      mixed6[0][0]
__________________________________________________________________________________________________
conv2d_63 (Conv2D)              (None, 17, 17, 192)  258048      activation_62[0][0]
__________________________________________________________________________________________________
conv2d_68 (Conv2D)              (None, 17, 17, 192)  258048      activation_67[0][0]
__________________________________________________________________________________________________
conv2d_69 (Conv2D)              (None, 17, 17, 192)  147456      average_pooling2d_6[0][0]
__________________________________________________________________________________________________
batch_normalization_60 (BatchNo (None, 17, 17, 192)  576         conv2d_60[0][0]
__________________________________________________________________________________________________
batch_normalization_63 (BatchNo (None, 17, 17, 192)  576         conv2d_63[0][0]
__________________________________________________________________________________________________
batch_normalization_68 (BatchNo (None, 17, 17, 192)  576         conv2d_68[0][0]
__________________________________________________________________________________________________
batch_normalization_69 (BatchNo (None, 17, 17, 192)  576         conv2d_69[0][0]
__________________________________________________________________________________________________
activation_60 (Activation)      (None, 17, 17, 192)  0           batch_normalization_60[0][0]
__________________________________________________________________________________________________
activation_63 (Activation)      (None, 17, 17, 192)  0           batch_normalization_63[0][0]
__________________________________________________________________________________________________
activation_68 (Activation)      (None, 17, 17, 192)  0           batch_normalization_68[0][0]
__________________________________________________________________________________________________
activation_69 (Activation)      (None, 17, 17, 192)  0           batch_normalization_69[0][0]
__________________________________________________________________________________________________
mixed7 (Concatenate)            (None, 17, 17, 768)  0           activation_60[0][0]
                                                                 activation_63[0][0]
                                                                 activation_68[0][0]
                                                                 activation_69[0][0]
__________________________________________________________________________________________________
conv2d_72 (Conv2D)              (None, 17, 17, 192)  147456      mixed7[0][0]
__________________________________________________________________________________________________
batch_normalization_72 (BatchNo (None, 17, 17, 192)  576         conv2d_72[0][0]
```

```
activation_72 (Activation)      (None, 17, 17, 192)  0           batch_normalization_72[0][0]
__________________________________________________________________________________________________
conv2d_73 (Conv2D)              (None, 17, 17, 192)  258048      activation_72[0][0]
__________________________________________________________________________________________________
batch_normalization_73 (BatchNo (None, 17, 17, 192)  576         conv2d_73[0][0]
__________________________________________________________________________________________________
activation_73 (Activation)      (None, 17, 17, 192)  0           batch_normalization_73[0][0]
__________________________________________________________________________________________________
conv2d_70 (Conv2D)              (None, 17, 17, 192)  147456      mixed7[0][0]
__________________________________________________________________________________________________
conv2d_74 (Conv2D)              (None, 17, 17, 192)  258048      activation_73[0][0]
__________________________________________________________________________________________________
batch_normalization_70 (BatchNo (None, 17, 17, 192)  576         conv2d_70[0][0]
__________________________________________________________________________________________________
batch_normalization_74 (BatchNo (None, 17, 17, 192)  576         conv2d_74[0][0]
__________________________________________________________________________________________________
activation_70 (Activation)      (None, 17, 17, 192)  0           batch_normalization_70[0][0]
__________________________________________________________________________________________________
activation_74 (Activation)      (None, 17, 17, 192)  0           batch_normalization_74[0][0]
__________________________________________________________________________________________________
conv2d_71 (Conv2D)              (None, 8, 8, 320)    552960      activation_70[0][0]
__________________________________________________________________________________________________
conv2d_75 (Conv2D)              (None, 8, 8, 192)    331776      activation_74[0][0]
__________________________________________________________________________________________________
batch_normalization_71 (BatchNo (None, 8, 8, 320)    960         conv2d_71[0][0]
__________________________________________________________________________________________________
batch_normalization_75 (BatchNo (None, 8, 8, 192)    576         conv2d_75[0][0]
__________________________________________________________________________________________________
activation_71 (Activation)      (None, 8, 8, 320)    0           batch_normalization_71[0][0]
__________________________________________________________________________________________________
activation_75 (Activation)      (None, 8, 8, 192)    0           batch_normalization_75[0][0]
__________________________________________________________________________________________________
max_pooling2d_3 (MaxPooling2D)  (None, 8, 8, 768)    0           mixed7[0][0]
__________________________________________________________________________________________________
mixed8 (Concatenate)            (None, 8, 8, 1280)   0           activation_71[0][0]
                                                                 activation_75[0][0]
                                                                 max_pooling2d_3[0][0]
__________________________________________________________________________________________________
conv2d_80 (Conv2D)              (None, 8, 8, 448)    573440      mixed8[0][0]
__________________________________________________________________________________________________
batch_normalization_80 (BatchNo (None, 8, 8, 448)    1344        conv2d_80[0][0]
__________________________________________________________________________________________________
activation_80 (Activation)      (None, 8, 8, 448)    0           batch_normalization_80[0][0]
__________________________________________________________________________________________________
conv2d_77 (Conv2D)              (None, 8, 8, 384)    491520      mixed8[0][0]
__________________________________________________________________________________________________
conv2d_81 (Conv2D)              (None, 8, 8, 384)    1548288     activation_80[0][0]
__________________________________________________________________________________________________
batch_normalization_77 (BatchNo (None, 8, 8, 384)    1152        conv2d_77[0][0]
__________________________________________________________________________________________________
batch_normalization_81 (BatchNo (None, 8, 8, 384)    1152        conv2d_81[0][0]
__________________________________________________________________________________________________
activation_77 (Activation)      (None, 8, 8, 384)    0           batch_normalization_77[0][0]
__________________________________________________________________________________________________
activation_81 (Activation)      (None, 8, 8, 384)    0           batch_normalization_81[0][0]
__________________________________________________________________________________________________
conv2d_78 (Conv2D)              (None, 8, 8, 384)    442368      activation_77[0][0]
__________________________________________________________________________________________________
conv2d_79 (Conv2D)              (None, 8, 8, 384)    442368      activation_77[0][0]
__________________________________________________________________________________________________
conv2d_82 (Conv2D)              (None, 8, 8, 384)    442368      activation_81[0][0]
__________________________________________________________________________________________________
conv2d_83 (Conv2D)              (None, 8, 8, 384)    442368      activation_81[0][0]
__________________________________________________________________________________________________
average_pooling2d_7 (AveragePoo (None, 8, 8, 1280)   0           mixed8[0][0]
__________________________________________________________________________________________________
conv2d_76 (Conv2D)              (None, 8, 8, 320)    409600      mixed8[0][0]
__________________________________________________________________________________________________
batch_normalization_78 (BatchNo (None, 8, 8, 384)    1152        conv2d_78[0][0]
__________________________________________________________________________________________________
batch_normalization_79 (BatchNo (None, 8, 8, 384)    1152        conv2d_79[0][0]
__________________________________________________________________________________________________
batch_normalization_82 (BatchNo (None, 8, 8, 384)    1152        conv2d_82[0][0]
__________________________________________________________________________________________________
batch_normalization_83 (BatchNo (None, 8, 8, 384)    1152        conv2d_83[0][0]
__________________________________________________________________________________________________
conv2d_84 (Conv2D)              (None, 8, 8, 192)    245760      average_pooling2d_7[0][0]
__________________________________________________________________________________________________
batch_normalization_76 (BatchNo (None, 8, 8, 320)    960         conv2d_76[0][0]
__________________________________________________________________________________________________
activation_78 (Activation)      (None, 8, 8, 384)    0           batch_normalization_78[0][0]
__________________________________________________________________________________________________
activation_79 (Activation)      (None, 8, 8, 384)    0           batch_normalization_79[0][0]
__________________________________________________________________________________________________
activation_82 (Activation)      (None, 8, 8, 384)    0           batch_normalization_82[0][0]
__________________________________________________________________________________________________
activation_83 (Activation)      (None, 8, 8, 384)    0           batch_normalization_83[0][0]
```

```
batch_normalization_84 (BatchNo (None, 8, 8, 192)    576         conv2d_84[0][0]
__________________________________________________________________________________________________
activation_76 (Activation)      (None, 8, 8, 320)    0           batch_normalization_76[0][0]
__________________________________________________________________________________________________
mixed9_0 (Concatenate)          (None, 8, 8, 768)    0           activation_78[0][0]
                                                                 activation_79[0][0]
__________________________________________________________________________________________________
concatenate (Concatenate)       (None, 8, 8, 768)    0           activation_82[0][0]
                                                                 activation_83[0][0]
__________________________________________________________________________________________________
activation_84 (Activation)      (None, 8, 8, 192)    0           batch_normalization_84[0][0]
__________________________________________________________________________________________________
mixed9 (Concatenate)            (None, 8, 8, 2048)   0           activation_76[0][0]
                                                                 mixed9_0[0][0]
                                                                 concatenate[0][0]
                                                                 activation_84[0][0]
__________________________________________________________________________________________________
conv2d_89 (Conv2D)              (None, 8, 8, 448)    917504      mixed9[0][0]
__________________________________________________________________________________________________
batch_normalization_89 (BatchNo (None, 8, 8, 448)    1344        conv2d_89[0][0]
__________________________________________________________________________________________________
activation_89 (Activation)      (None, 8, 8, 448)    0           batch_normalization_89[0][0]
__________________________________________________________________________________________________
conv2d_86 (Conv2D)              (None, 8, 8, 384)    786432      mixed9[0][0]
__________________________________________________________________________________________________
conv2d_90 (Conv2D)              (None, 8, 8, 384)    1548288     activation_89[0][0]
__________________________________________________________________________________________________
batch_normalization_86 (BatchNo (None, 8, 8, 384)    1152        conv2d_86[0][0]
__________________________________________________________________________________________________
batch_normalization_90 (BatchNo (None, 8, 8, 384)    1152        conv2d_90[0][0]
__________________________________________________________________________________________________
activation_86 (Activation)      (None, 8, 8, 384)    0           batch_normalization_86[0][0]
__________________________________________________________________________________________________
activation_90 (Activation)      (None, 8, 8, 384)    0           batch_normalization_90[0][0]
__________________________________________________________________________________________________
conv2d_87 (Conv2D)              (None, 8, 8, 384)    442368      activation_86[0][0]
__________________________________________________________________________________________________
conv2d_88 (Conv2D)              (None, 8, 8, 384)    442368      activation_86[0][0]
__________________________________________________________________________________________________
conv2d_91 (Conv2D)              (None, 8, 8, 384)    442368      activation_90[0][0]
__________________________________________________________________________________________________
conv2d_92 (Conv2D)              (None, 8, 8, 384)    442368      activation_90[0][0]
__________________________________________________________________________________________________
average_pooling2d_8 (AveragePoo (None, 8, 8, 2048)   0           mixed9[0][0]
__________________________________________________________________________________________________
conv2d_85 (Conv2D)              (None, 8, 8, 320)    655360      mixed9[0][0]
__________________________________________________________________________________________________
batch_normalization_87 (BatchNo (None, 8, 8, 384)    1152        conv2d_87[0][0]
__________________________________________________________________________________________________
batch_normalization_88 (BatchNo (None, 8, 8, 384)    1152        conv2d_88[0][0]
__________________________________________________________________________________________________
batch_normalization_91 (BatchNo (None, 8, 8, 384)    1152        conv2d_91[0][0]
__________________________________________________________________________________________________
batch_normalization_92 (BatchNo (None, 8, 8, 384)    1152        conv2d_92[0][0]
__________________________________________________________________________________________________
conv2d_93 (Conv2D)              (None, 8, 8, 192)    393216      average_pooling2d_8[0][0]
__________________________________________________________________________________________________
batch_normalization_85 (BatchNo (None, 8, 8, 320)    960         conv2d_85[0][0]
__________________________________________________________________________________________________
activation_87 (Activation)      (None, 8, 8, 384)    0           batch_normalization_87[0][0]
__________________________________________________________________________________________________
activation_88 (Activation)      (None, 8, 8, 384)    0           batch_normalization_88[0][0]
__________________________________________________________________________________________________
activation_91 (Activation)      (None, 8, 8, 384)    0           batch_normalization_91[0][0]
__________________________________________________________________________________________________
activation_92 (Activation)      (None, 8, 8, 384)    0           batch_normalization_92[0][0]
__________________________________________________________________________________________________
batch_normalization_93 (BatchNo (None, 8, 8, 192)    576         conv2d_93[0][0]
__________________________________________________________________________________________________
activation_85 (Activation)      (None, 8, 8, 320)    0           batch_normalization_85[0][0]
__________________________________________________________________________________________________
mixed9_1 (Concatenate)          (None, 8, 8, 768)    0           activation_87[0][0]
                                                                 activation_88[0][0]
__________________________________________________________________________________________________
concatenate_1 (Concatenate)     (None, 8, 8, 768)    0           activation_91[0][0]
                                                                 activation_92[0][0]
__________________________________________________________________________________________________
activation_93 (Activation)      (None, 8, 8, 192)    0           batch_normalization_93[0][0]
__________________________________________________________________________________________________
mixed10 (Concatenate)           (None, 8, 8, 2048)   0           activation_85[0][0]
                                                                 mixed9_1[0][0]
                                                                 concatenate_1[0][0]
                                                                 activation_93[0][0]
__________________________________________________________________________________________________
avg_pool (GlobalAveragePooling2 (None, 2048)         0           mixed10[0][0]
__________________________________________________________________________________________________
predictions (Dense)             (None, 1000)         2049000     avg_pool[0][0]
==================================================================================================
```

```
Total params: 23,851,784
Trainable params: 23,817,352
Non-trainable params: 34,432
```

Appendix B: ResNet 50 V2 Model Summary

```
Model: "resnet50v2"
```

Layer (type)	Output Shape	Param #	Connected to
input_1 (InputLayer)	[(None, 224, 224, 3)	0	
conv1_pad (ZeroPadding2D)	(None, 230, 230, 3)	0	input_1[0][0]
conv1_conv (Conv2D)	(None, 112, 112, 64)	9472	conv1_pad[0][0]
pool1_pad (ZeroPadding2D)	(None, 114, 114, 64)	0	conv1_conv[0][0]
pool1_pool (MaxPooling2D)	(None, 56, 56, 64)	0	pool1_pad[0][0]
conv2_block1_preact_bn (BatchNo	(None, 56, 56, 64)	256	pool1_pool[0][0]
conv2_block1_preact_relu (Activ	(None, 56, 56, 64)	0	conv2_block1_preact_bn[0][0]
conv2_block1_1_conv (Conv2D)	(None, 56, 56, 64)	4096	conv2_block1_preact_relu[0][0]
conv2_block1_1_bn (BatchNormali	(None, 56, 56, 64)	256	conv2_block1_1_conv[0][0]
conv2_block1_1_relu (Activation	(None, 56, 56, 64)	0	conv2_block1_1_bn[0][0]
conv2_block1_2_pad (ZeroPadding	(None, 58, 58, 64)	0	conv2_block1_1_relu[0][0]
conv2_block1_2_conv (Conv2D)	(None, 56, 56, 64)	36864	conv2_block1_2_pad[0][0]
conv2_block1_2_bn (BatchNormali	(None, 56, 56, 64)	256	conv2_block1_2_conv[0][0]
conv2_block1_2_relu (Activation	(None, 56, 56, 64)	0	conv2_block1_2_bn[0][0]
conv2_block1_0_conv (Conv2D)	(None, 56, 56, 256)	16640	conv2_block1_preact_relu[0][0]
conv2_block1_3_conv (Conv2D)	(None, 56, 56, 256)	16640	conv2_block1_2_relu[0][0]
conv2_block1_out (Add)	(None, 56, 56, 256)	0	conv2_block1_0_conv[0][0] conv2_block1_3_conv[0][0]
conv2_block2_preact_bn (BatchNo	(None, 56, 56, 256)	1024	conv2_block1_out[0][0]
conv2_block2_preact_relu (Activ	(None, 56, 56, 256)	0	conv2_block2_preact_bn[0][0]
conv2_block2_1_conv (Conv2D)	(None, 56, 56, 64)	16384	conv2_block2_preact_relu[0][0]
conv2_block2_1_bn (BatchNormali	(None, 56, 56, 64)	256	conv2_block2_1_conv[0][0]
conv2_block2_1_relu (Activation	(None, 56, 56, 64)	0	conv2_block2_1_bn[0][0]
conv2_block2_2_pad (ZeroPadding	(None, 58, 58, 64)	0	conv2_block2_1_relu[0][0]
conv2_block2_2_conv (Conv2D)	(None, 56, 56, 64)	36864	conv2_block2_2_pad[0][0]
conv2_block2_2_bn (BatchNormali	(None, 56, 56, 64)	256	conv2_block2_2_conv[0][0]
conv2_block2_2_relu (Activation	(None, 56, 56, 64)	0	conv2_block2_2_bn[0][0]
conv2_block2_3_conv (Conv2D)	(None, 56, 56, 256)	16640	conv2_block2_2_relu[0][0]
conv2_block2_out (Add)	(None, 56, 56, 256)	0	conv2_block1_out[0][0] conv2_block2_3_conv[0][0]
conv2_block3_preact_bn (BatchNo	(None, 56, 56, 256)	1024	conv2_block2_out[0][0]
conv2_block3_preact_relu (Activ	(None, 56, 56, 256)	0	conv2_block3_preact_bn[0][0]
conv2_block3_1_conv (Conv2D)	(None, 56, 56, 64)	16384	conv2_block3_preact_relu[0][0]
conv2_block3_1_bn (BatchNormali	(None, 56, 56, 64)	256	conv2_block3_1_conv[0][0]
conv2_block3_1_relu (Activation	(None, 56, 56, 64)	0	conv2_block3_1_bn[0][0]
conv2_block3_2_pad (ZeroPadding	(None, 58, 58, 64)	0	conv2_block3_1_relu[0][0]
conv2_block3_2_conv (Conv2D)	(None, 28, 28, 64)	36864	conv2_block3_2_pad[0][0]

```
conv2_block3_2_bn (BatchNormali (None, 28, 28, 64)   256         conv2_block3_2_conv[0][0]
conv2_block3_2_relu (Activation (None, 28, 28, 64)   0           conv2_block3_2_bn[0][0]
max_pooling2d (MaxPooling2D)    (None, 28, 28, 256)  0           conv2_block2_out[0][0]
conv2_block3_3_conv (Conv2D)    (None, 28, 28, 256)  16640       conv2_block3_2_relu[0][0]
conv2_block3_out (Add)          (None, 28, 28, 256)  0           max_pooling2d[0][0]
                                                                 conv2_block3_3_conv[0][0]
conv3_block1_preact_bn (BatchNo (None, 28, 28, 256)  1024        conv2_block3_out[0][0]
conv3_block1_preact_relu (Activ (None, 28, 28, 256)  0           conv3_block1_preact_bn[0][0]
conv3_block1_1_conv (Conv2D)    (None, 28, 28, 128)  32768       conv3_block1_preact_relu[0][0]
conv3_block1_1_bn (BatchNormali (None, 28, 28, 128)  512         conv3_block1_1_conv[0][0]
conv3_block1_1_relu (Activation (None, 28, 28, 128)  0           conv3_block1_1_bn[0][0]
conv3_block1_2_pad (ZeroPadding (None, 30, 30, 128)  0           conv3_block1_1_relu[0][0]
conv3_block1_2_conv (Conv2D)    (None, 28, 28, 128)  147456      conv3_block1_2_pad[0][0]
conv3_block1_2_bn (BatchNormali (None, 28, 28, 128)  512         conv3_block1_2_conv[0][0]
conv3_block1_2_relu (Activation (None, 28, 28, 128)  0           conv3_block1_2_bn[0][0]
conv3_block1_0_conv (Conv2D)    (None, 28, 28, 512)  131584      conv3_block1_preact_relu[0][0]
conv3_block1_3_conv (Conv2D)    (None, 28, 28, 512)  66048       conv3_block1_2_relu[0][0]
conv3_block1_out (Add)          (None, 28, 28, 512)  0           conv3_block1_0_conv[0][0]
                                                                 conv3_block1_3_conv[0][0]
conv3_block2_preact_bn (BatchNo (None, 28, 28, 512)  2048        conv3_block1_out[0][0]
conv3_block2_preact_relu (Activ (None, 28, 28, 512)  0           conv3_block2_preact_bn[0][0]
conv3_block2_1_conv (Conv2D)    (None, 28, 28, 128)  65536       conv3_block2_preact_relu[0][0]
conv3_block2_1_bn (BatchNormali (None, 28, 28, 128)  512         conv3_block2_1_conv[0][0]
conv3_block2_1_relu (Activation (None, 28, 28, 128)  0           conv3_block2_1_bn[0][0]
conv3_block2_2_pad (ZeroPadding (None, 30, 30, 128)  0           conv3_block2_1_relu[0][0]
conv3_block2_2_conv (Conv2D)    (None, 28, 28, 128)  147456      conv3_block2_2_pad[0][0]
conv3_block2_2_bn (BatchNormali (None, 28, 28, 128)  512         conv3_block2_2_conv[0][0]
conv3_block2_2_relu (Activation (None, 28, 28, 128)  0           conv3_block2_2_bn[0][0]
conv3_block2_3_conv (Conv2D)    (None, 28, 28, 512)  66048       conv3_block2_2_relu[0][0]
conv3_block2_out (Add)          (None, 28, 28, 512)  0           conv3_block1_out[0][0]
                                                                 conv3_block2_3_conv[0][0]
conv3_block3_preact_bn (BatchNo (None, 28, 28, 512)  2048        conv3_block2_out[0][0]
conv3_block3_preact_relu (Activ (None, 28, 28, 512)  0           conv3_block3_preact_bn[0][0]
conv3_block3_1_conv (Conv2D)    (None, 28, 28, 128)  65536       conv3_block3_preact_relu[0][0]
conv3_block3_1_bn (BatchNormali (None, 28, 28, 128)  512         conv3_block3_1_conv[0][0]
conv3_block3_1_relu (Activation (None, 28, 28, 128)  0           conv3_block3_1_bn[0][0]
conv3_block3_2_pad (ZeroPadding (None, 30, 30, 128)  0           conv3_block3_1_relu[0][0]
conv3_block3_2_conv (Conv2D)    (None, 28, 28, 128)  147456      conv3_block3_2_pad[0][0]
conv3_block3_2_bn (BatchNormali (None, 28, 28, 128)  512         conv3_block3_2_conv[0][0]
conv3_block3_2_relu (Activation (None, 28, 28, 128)  0           conv3_block3_2_bn[0][0]
conv3_block3_3_conv (Conv2D)    (None, 28, 28, 512)  66048       conv3_block3_2_relu[0][0]
conv3_block3_out (Add)          (None, 28, 28, 512)  0           conv3_block2_out[0][0]
                                                                 conv3_block3_3_conv[0][0]
conv3_block4_preact_bn (BatchNo (None, 28, 28, 512)  2048        conv3_block3_out[0][0]
conv3_block4_preact_relu (Activ (None, 28, 28, 512)  0           conv3_block4_preact_bn[0][0]
conv3_block4_1_conv (Conv2D)    (None, 28, 28, 128)  65536       conv3_block4_preact_relu[0][0]
```

```
conv3_block4_1_bn (BatchNormali (None, 28, 28, 128)  512         conv3_block4_1_conv[0][0]
conv3_block4_1_relu (Activation (None, 28, 28, 128)  0           conv3_block4_1_bn[0][0]
conv3_block4_2_pad (ZeroPadding (None, 30, 30, 128)  0           conv3_block4_1_relu[0][0]
conv3_block4_2_conv (Conv2D)    (None, 14, 14, 128)  147456      conv3_block4_2_pad[0][0]
conv3_block4_2_bn (BatchNormali (None, 14, 14, 128)  512         conv3_block4_2_conv[0][0]
conv3_block4_2_relu (Activation (None, 14, 14, 128)  0           conv3_block4_2_bn[0][0]
max_pooling2d_1 (MaxPooling2D)  (None, 14, 14, 512)  0           conv3_block3_out[0][0]
conv3_block4_3_conv (Conv2D)    (None, 14, 14, 512)  66048       conv3_block4_2_relu[0][0]
conv3_block4_out (Add)          (None, 14, 14, 512)  0           max_pooling2d_1[0][0]
                                                                 conv3_block4_3_conv[0][0]
conv4_block1_preact_bn (BatchNo (None, 14, 14, 512)  2048        conv3_block4_out[0][0]
conv4_block1_preact_relu (Activ (None, 14, 14, 512)  0           conv4_block1_preact_bn[0][0]
conv4_block1_1_conv (Conv2D)    (None, 14, 14, 256)  131072      conv4_block1_preact_relu[0][0]
conv4_block1_1_bn (BatchNormali (None, 14, 14, 256)  1024        conv4_block1_1_conv[0][0]
conv4_block1_1_relu (Activation (None, 14, 14, 256)  0           conv4_block1_1_bn[0][0]
conv4_block1_2_pad (ZeroPadding (None, 16, 16, 256)  0           conv4_block1_1_relu[0][0]
conv4_block1_2_conv (Conv2D)    (None, 14, 14, 256)  589824      conv4_block1_2_pad[0][0]
conv4_block1_2_bn (BatchNormali (None, 14, 14, 256)  1024        conv4_block1_2_conv[0][0]
conv4_block1_2_relu (Activation (None, 14, 14, 256)  0           conv4_block1_2_bn[0][0]
conv4_block1_0_conv (Conv2D)    (None, 14, 14, 1024) 525312      conv4_block1_preact_relu[0][0]
conv4_block1_3_conv (Conv2D)    (None, 14, 14, 1024) 263168      conv4_block1_2_relu[0][0]
conv4_block1_out (Add)          (None, 14, 14, 1024) 0           conv4_block1_0_conv[0][0]
                                                                 conv4_block1_3_conv[0][0]
conv4_block2_preact_bn (BatchNo (None, 14, 14, 1024) 4096        conv4_block1_out[0][0]
conv4_block2_preact_relu (Activ (None, 14, 14, 1024) 0           conv4_block2_preact_bn[0][0]
conv4_block2_1_conv (Conv2D)    (None, 14, 14, 256)  262144      conv4_block2_preact_relu[0][0]
conv4_block2_1_bn (BatchNormali (None, 14, 14, 256)  1024        conv4_block2_1_conv[0][0]
conv4_block2_1_relu (Activation (None, 14, 14, 256)  0           conv4_block2_1_bn[0][0]
conv4_block2_2_pad (ZeroPadding (None, 16, 16, 256)  0           conv4_block2_1_relu[0][0]
conv4_block2_2_conv (Conv2D)    (None, 14, 14, 256)  589824      conv4_block2_2_pad[0][0]
conv4_block2_2_bn (BatchNormali (None, 14, 14, 256)  1024        conv4_block2_2_conv[0][0]
conv4_block2_2_relu (Activation (None, 14, 14, 256)  0           conv4_block2_2_bn[0][0]
conv4_block2_3_conv (Conv2D)    (None, 14, 14, 1024) 263168      conv4_block2_2_relu[0][0]
conv4_block2_out (Add)          (None, 14, 14, 1024) 0           conv4_block1_out[0][0]
                                                                 conv4_block2_3_conv[0][0]
conv4_block3_preact_bn (BatchNo (None, 14, 14, 1024) 4096        conv4_block2_out[0][0]
conv4_block3_preact_relu (Activ (None, 14, 14, 1024) 0           conv4_block3_preact_bn[0][0]
conv4_block3_1_conv (Conv2D)    (None, 14, 14, 256)  262144      conv4_block3_preact_relu[0][0]
conv4_block3_1_bn (BatchNormali (None, 14, 14, 256)  1024        conv4_block3_1_conv[0][0]
conv4_block3_1_relu (Activation (None, 14, 14, 256)  0           conv4_block3_1_bn[0][0]
conv4_block3_2_pad (ZeroPadding (None, 16, 16, 256)  0           conv4_block3_1_relu[0][0]
conv4_block3_2_conv (Conv2D)    (None, 14, 14, 256)  589824      conv4_block3_2_pad[0][0]
conv4_block3_2_bn (BatchNormali (None, 14, 14, 256)  1024        conv4_block3_2_conv[0][0]
conv4_block3_2_relu (Activation (None, 14, 14, 256)  0           conv4_block3_2_bn[0][0]
conv4_block3_3_conv (Conv2D)    (None, 14, 14, 1024) 263168      conv4_block3_2_relu[0][0]
```

```
conv4_block3_out (Add)          (None, 14, 14, 1024) 0           conv4_block2_out[0][0]
                                                                 conv4_block3_3_conv[0][0]
__________________________________________________________________________________________
conv4_block4_preact_bn (BatchNo (None, 14, 14, 1024) 4096        conv4_block3_out[0][0]
__________________________________________________________________________________________
conv4_block4_preact_relu (Activ (None, 14, 14, 1024) 0           conv4_block4_preact_bn[0][0]
__________________________________________________________________________________________
conv4_block4_1_conv (Conv2D)    (None, 14, 14, 256)  262144      conv4_block4_preact_relu[0][0]
__________________________________________________________________________________________
conv4_block4_1_bn (BatchNormali (None, 14, 14, 256)  1024        conv4_block4_1_conv[0][0]
__________________________________________________________________________________________
conv4_block4_1_relu (Activation (None, 14, 14, 256)  0           conv4_block4_1_bn[0][0]
__________________________________________________________________________________________
conv4_block4_2_pad (ZeroPadding (None, 16, 16, 256)  0           conv4_block4_1_relu[0][0]
__________________________________________________________________________________________
conv4_block4_2_conv (Conv2D)    (None, 14, 14, 256)  589824      conv4_block4_2_pad[0][0]
__________________________________________________________________________________________
conv4_block4_2_bn (BatchNormali (None, 14, 14, 256)  1024        conv4_block4_2_conv[0][0]
__________________________________________________________________________________________
conv4_block4_2_relu (Activation (None, 14, 14, 256)  0           conv4_block4_2_bn[0][0]
__________________________________________________________________________________________
conv4_block4_3_conv (Conv2D)    (None, 14, 14, 1024) 263168      conv4_block4_2_relu[0][0]
__________________________________________________________________________________________
conv4_block4_out (Add)          (None, 14, 14, 1024) 0           conv4_block3_out[0][0]
                                                                 conv4_block4_3_conv[0][0]
__________________________________________________________________________________________
conv4_block5_preact_bn (BatchNo (None, 14, 14, 1024) 4096        conv4_block4_out[0][0]
__________________________________________________________________________________________
conv4_block5_preact_relu (Activ (None, 14, 14, 1024) 0           conv4_block5_preact_bn[0][0]
__________________________________________________________________________________________
conv4_block5_1_conv (Conv2D)    (None, 14, 14, 256)  262144      conv4_block5_preact_relu[0][0]
__________________________________________________________________________________________
conv4_block5_1_bn (BatchNormali (None, 14, 14, 256)  1024        conv4_block5_1_conv[0][0]
__________________________________________________________________________________________
conv4_block5_1_relu (Activation (None, 14, 14, 256)  0           conv4_block5_1_bn[0][0]
__________________________________________________________________________________________
conv4_block5_2_pad (ZeroPadding (None, 16, 16, 256)  0           conv4_block5_1_relu[0][0]
__________________________________________________________________________________________
conv4_block5_2_conv (Conv2D)    (None, 14, 14, 256)  589824      conv4_block5_2_pad[0][0]
__________________________________________________________________________________________
conv4_block5_2_bn (BatchNormali (None, 14, 14, 256)  1024        conv4_block5_2_conv[0][0]
__________________________________________________________________________________________
conv4_block5_2_relu (Activation (None, 14, 14, 256)  0           conv4_block5_2_bn[0][0]
__________________________________________________________________________________________
conv4_block5_3_conv (Conv2D)    (None, 14, 14, 1024) 263168      conv4_block5_2_relu[0][0]
__________________________________________________________________________________________
conv4_block5_out (Add)          (None, 14, 14, 1024) 0           conv4_block4_out[0][0]
                                                                 conv4_block5_3_conv[0][0]
__________________________________________________________________________________________
conv4_block6_preact_bn (BatchNo (None, 14, 14, 1024) 4096        conv4_block5_out[0][0]
__________________________________________________________________________________________
conv4_block6_preact_relu (Activ (None, 14, 14, 1024) 0           conv4_block6_preact_bn[0][0]
__________________________________________________________________________________________
conv4_block6_1_conv (Conv2D)    (None, 14, 14, 256)  262144      conv4_block6_preact_relu[0][0]
__________________________________________________________________________________________
conv4_block6_1_bn (BatchNormali (None, 14, 14, 256)  1024        conv4_block6_1_conv[0][0]
__________________________________________________________________________________________
conv4_block6_1_relu (Activation (None, 14, 14, 256)  0           conv4_block6_1_bn[0][0]
__________________________________________________________________________________________
conv4_block6_2_pad (ZeroPadding (None, 16, 16, 256)  0           conv4_block6_1_relu[0][0]
__________________________________________________________________________________________
conv4_block6_2_conv (Conv2D)    (None, 7, 7, 256)    589824      conv4_block6_2_pad[0][0]
__________________________________________________________________________________________
conv4_block6_2_bn (BatchNormali (None, 7, 7, 256)    1024        conv4_block6_2_conv[0][0]
__________________________________________________________________________________________
conv4_block6_2_relu (Activation (None, 7, 7, 256)    0           conv4_block6_2_bn[0][0]
__________________________________________________________________________________________
max_pooling2d_2 (MaxPooling2D)  (None, 7, 7, 1024)   0           conv4_block5_out[0][0]
__________________________________________________________________________________________
conv4_block6_3_conv (Conv2D)    (None, 7, 7, 1024)   263168      conv4_block6_2_relu[0][0]
__________________________________________________________________________________________
conv4_block6_out (Add)          (None, 7, 7, 1024)   0           max_pooling2d_2[0][0]
                                                                 conv4_block6_3_conv[0][0]
__________________________________________________________________________________________
conv5_block1_preact_bn (BatchNo (None, 7, 7, 1024)   4096        conv4_block6_out[0][0]
__________________________________________________________________________________________
conv5_block1_preact_relu (Activ (None, 7, 7, 1024)   0           conv5_block1_preact_bn[0][0]
__________________________________________________________________________________________
conv5_block1_1_conv (Conv2D)    (None, 7, 7, 512)    524288      conv5_block1_preact_relu[0][0]
__________________________________________________________________________________________
conv5_block1_1_bn (BatchNormali (None, 7, 7, 512)    2048        conv5_block1_1_conv[0][0]
__________________________________________________________________________________________
conv5_block1_1_relu (Activation (None, 7, 7, 512)    0           conv5_block1_1_bn[0][0]
__________________________________________________________________________________________
conv5_block1_2_pad (ZeroPadding (None, 9, 9, 512)    0           conv5_block1_1_relu[0][0]
__________________________________________________________________________________________
conv5_block1_2_conv (Conv2D)    (None, 7, 7, 512)    2359296     conv5_block1_2_pad[0][0]
```

```
conv5_block1_2_bn (BatchNormali (None, 7, 7, 512)    2048        conv5_block1_2_conv[0][0]
__________________________________________________________________________________________________
conv5_block1_2_relu (Activation (None, 7, 7, 512)    0           conv5_block1_2_bn[0][0]
__________________________________________________________________________________________________
conv5_block1_0_conv (Conv2D)    (None, 7, 7, 2048)   2099200     conv5_block1_preact_relu[0][0]
__________________________________________________________________________________________________
conv5_block1_3_conv (Conv2D)    (None, 7, 7, 2048)   1050624     conv5_block1_2_relu[0][0]
__________________________________________________________________________________________________
conv5_block1_out (Add)          (None, 7, 7, 2048)   0           conv5_block1_0_conv[0][0]
                                                                 conv5_block1_3_conv[0][0]
__________________________________________________________________________________________________
conv5_block2_preact_bn (BatchNo (None, 7, 7, 2048)   8192        conv5_block1_out[0][0]
__________________________________________________________________________________________________
conv5_block2_preact_relu (Activ (None, 7, 7, 2048)   0           conv5_block2_preact_bn[0][0]
__________________________________________________________________________________________________
conv5_block2_1_conv (Conv2D)    (None, 7, 7, 512)    1048576     conv5_block2_preact_relu[0][0]
__________________________________________________________________________________________________
conv5_block2_1_bn (BatchNormali (None, 7, 7, 512)    2048        conv5_block2_1_conv[0][0]
__________________________________________________________________________________________________
conv5_block2_1_relu (Activation (None, 7, 7, 512)    0           conv5_block2_1_bn[0][0]
__________________________________________________________________________________________________
conv5_block2_2_pad (ZeroPadding (None, 9, 9, 512)    0           conv5_block2_1_relu[0][0]
__________________________________________________________________________________________________
conv5_block2_2_conv (Conv2D)    (None, 7, 7, 512)    2359296     conv5_block2_2_pad[0][0]
__________________________________________________________________________________________________
conv5_block2_2_bn (BatchNormali (None, 7, 7, 512)    2048        conv5_block2_2_conv[0][0]
__________________________________________________________________________________________________
conv5_block2_2_relu (Activation (None, 7, 7, 512)    0           conv5_block2_2_bn[0][0]
__________________________________________________________________________________________________
conv5_block2_3_conv (Conv2D)    (None, 7, 7, 2048)   1050624     conv5_block2_2_relu[0][0]
__________________________________________________________________________________________________
conv5_block2_out (Add)          (None, 7, 7, 2048)   0           conv5_block1_out[0][0]
                                                                 conv5_block2_3_conv[0][0]
__________________________________________________________________________________________________
conv5_block3_preact_bn (BatchNo (None, 7, 7, 2048)   8192        conv5_block2_out[0][0]
__________________________________________________________________________________________________
conv5_block3_preact_relu (Activ (None, 7, 7, 2048)   0           conv5_block3_preact_bn[0][0]
__________________________________________________________________________________________________
conv5_block3_1_conv (Conv2D)    (None, 7, 7, 512)    1048576     conv5_block3_preact_relu[0][0]
__________________________________________________________________________________________________
conv5_block3_1_bn (BatchNormali (None, 7, 7, 512)    2048        conv5_block3_1_conv[0][0]
__________________________________________________________________________________________________
conv5_block3_1_relu (Activation (None, 7, 7, 512)    0           conv5_block3_1_bn[0][0]
__________________________________________________________________________________________________
conv5_block3_2_pad (ZeroPadding (None, 9, 9, 512)    0           conv5_block3_1_relu[0][0]
__________________________________________________________________________________________________
conv5_block3_2_conv (Conv2D)    (None, 7, 7, 512)    2359296     conv5_block3_2_pad[0][0]
__________________________________________________________________________________________________
conv5_block3_2_bn (BatchNormali (None, 7, 7, 512)    2048        conv5_block3_2_conv[0][0]
__________________________________________________________________________________________________
conv5_block3_2_relu (Activation (None, 7, 7, 512)    0           conv5_block3_2_bn[0][0]
__________________________________________________________________________________________________
conv5_block3_3_conv (Conv2D)    (None, 7, 7, 2048)   1050624     conv5_block3_2_relu[0][0]
__________________________________________________________________________________________________
conv5_block3_out (Add)          (None, 7, 7, 2048)   0           conv5_block2_out[0][0]
                                                                 conv5_block3_3_conv[0][0]
__________________________________________________________________________________________________
post_bn (BatchNormalization)    (None, 7, 7, 2048)   8192        conv5_block3_out[0][0]
__________________________________________________________________________________________________
post_relu (Activation)          (None, 7, 7, 2048)   0           post_bn[0][0]
__________________________________________________________________________________________________
avg_pool (GlobalAveragePooling2 (None, 2048)         0           post_relu[0][0]
__________________________________________________________________________________________________
predictions (Dense)             (None, 1000)         2049000     avg_pool[0][0]
==================================================================================================
Total params: 25,613,800
Trainable params: 25,568,360
Non-trainable params: 45,440
__________________________________________________________________________________________________
```