CS783 Assignment 1

Anubhav Mittal, Tanuj

February 5th, 2019

1 Approaches Used

1.1 Transfer Learning(Main Model)

Since the dataset provided to us is small in size (3500 images), using a pretrained network as a feature vector extractor is a viable approach to the problem. We used ResNet50 for extracting the feature vector of 2048 dimensions, and then trained a classifier on this for the 16 different classes possible using a softmax activation layer. For performing the task presented, we feed the test image to the network and receive the 16-dimension output. We then rank all the training images of the classes sorted in decreasing order of their score i.e. we place all the images of the class with the highest score at the top.

1.2 Selective Search + Transfer Learning

Since we have to deal with images containing multiple objects, before we move on to the stage of Object classification, we first need to identify the objects i.e Object Detection. So to learn the bounding boxes for objects, we looked at and tried Selective Search to identify objects. The idea was to get bounding boxes for different objects using selective search, feed each bounding box with rank better than a threshold to the neural net trained in section 1.1 and rank the classes based on the no. of bounding boxes assigned to each class. Then rank the training images similar to section 1.1. So for each query image, we would proceed in the following way: Run selective search algorithm to to identify high scoring bounding boxes and feed each of these bounding boxes to the Transfer Learning Model. Finally rank the classes in the order of the number of bounding boxes assigned to each class.

1.3 Bag of Visual words with ORB descriptors

We first looked at ORB (we used ORB as it is freely available) to obtain local features/ feature vector and then classify the query image via TF-IDF score as discussed in class. We first extracted the visual descriptors from each image in the training set. The cluster centers were obtained via the K-Means Clustering. Various values of number of cluster centers were tried and it maximized for k=160 .
We then represented each image by a historgram in the following way:first created a vector of k value for each image and then for each keypoints in an image, found the nearest center and increase by one its value, and finally normalised it. Representations for the query images were made similarly.
We then tried two different methods after getting the above-defined representations:

1.3.1 Cosine Product

For obtaining ranking of the training images, we calculated the cosine product between the query image and the training images representations, and then ranked the images in decreasing order of the cosine product values. The intuition behind was that images with common object will have a similar higher component in the vector embedding, leading to similar direction and greater cosine product.

1.3.2 Neural Network classifier

After obtaining the vector embedding, the idea was to train a classifier using a neural net, similar to the transfer learning approach, where we assign each image to a class and for the query image, rank all the training images of the classes sorted in decreasing order of their score from the neural net.

2 Results

2.1 Transfer Learning

The results from the transfer learning based model were the most promising, hence this is our main model.

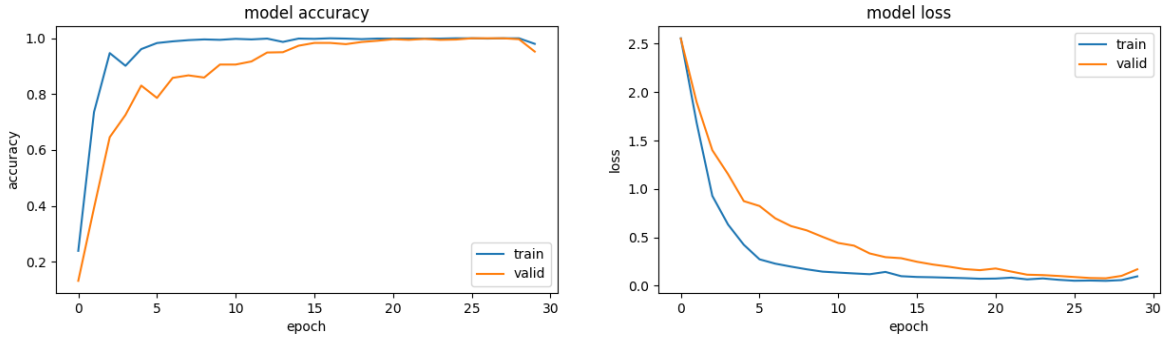

Figure 1: Accuracy and Loss vs iterations for training of the classifier layer

For the easy test cases, the model predicted correct objects most of the time, but failed in the hard cases. This is discussed in section 3.

2.2 Bag of Visual words with ORB descriptors

2.2.1 Neural Network classifier

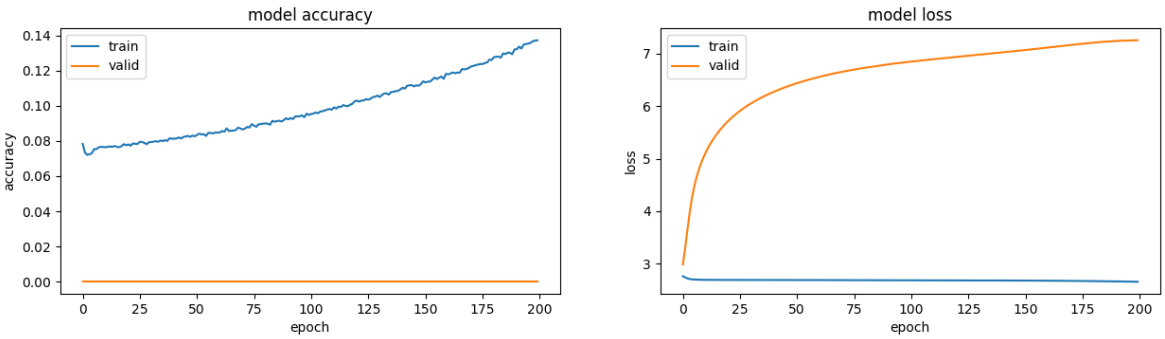

Figure 2: Accuracy and Loss vs iterations for training of the classifier

As the loss for validation set increases and accuracy remains zero, we concluded the model was not able to solve the problem at hand.

2.2.2 Cosine Product

The ranking was not uniform in this case, with many images whose objects are not in the query image making to the top of the rankings. Hence, the transfer learning approach was chosen over this.

3 Challenges Faced

Object detection was a very difficult task.Even in some of the highest scoring bounding box after selective search, majority of bounding box were filled with clutter. Hence, selective search method was not used.

It was observed that as in training we have images with single objects in them, the models perform fairly well on test queries containing single objects with a white background. However, for multiple objects, some of the present objects classes have high probability while others get a lower ranking.

Also, problems arise in case of difficult backgrounds. One example is that the listerine green images, which were having high ranking in the forest images. Due to this and the abundance of many edges and textures in the images, it is difficult to predict the object in the images.

4 Hashes for models and files saved

1. best.hdf5 - 7e00d682fc6e6c517732ab98516d54a0
2. BagOfWords_Features.npy - 2e3da0959f7dc8746f51e742a60199f5
3. BagOfWords_Labels.npy - 773d247a0681cf3aaa97a5287a75e21f
4. ImageName.npy - 741b63484124661aa38cbf57f12f2a88
5. Image_Name_resnet.npy - 4ec3446d1960d166a787b68b435b41fb
6. kmeans.pickle - 502af3aa234470d16415ca4a5ecce882