Homework Assignment 1

Abhay Gupta (Andrew Id: abhayg)

2018/09/18

1 Question 1

1.1 Extracting Filter Responses

1.1.1 Q1.1.1

1. **Gaussian Filter**: It is a smoothening low-pass filter - it removes high-frequencies from the image (details and noise) and allows lower frequencies to pass (blurring of the image)
2. **Laplacian of Gaussian Filter**: Calculates the second spatial derivative of the image, particularly it measures the change in intensities across the image regions - which means that when there are uniform regions in the image - the response of the filter applied to the image will be zero - however, in regions where intensities change, the LoG filter has positive values on the darker side and negative values on the lighter side. For an edge on the image, the response of the filter will be zero.
3. **Gaussian x-derivative**: This is used to capture the vertical edges in the image, when the filter is applied to the image.
4. **Gaussian y-derivative**: This is used to capture the horizontal edges in the image, when the filter is applied to the image.

Since all filters do use gaussian inherently, they all perform some level of smoothening to the images - and the different scales are used to decide the level of smoothening (blurring) of the image. The higher the value of σ is, the greater the blurring.

1.1.2 Q1.1.2

The image convolved with all 20 filters can be seen in Figure 1

1.2 Q1.2

Code is included in the submission.

1.3 Q1.3

Adding wordmap for three images, we get Figure 2.

We can see that the clustering produces semantic information in the image allowing for similar features of the image to be learned. For example, the visualization of the highway image makes it possible to learn near regions of the road (highway) with higher intensity and the green regions of the road (highway) with lower intensities for the image, giving more semantic information with respect to the perception of the image. For the windmill image, we can see that deeper blue colors in the sky have a different wordmap representation that the lighter sections on the sky - giving more information on the illumination aspects of the image too.

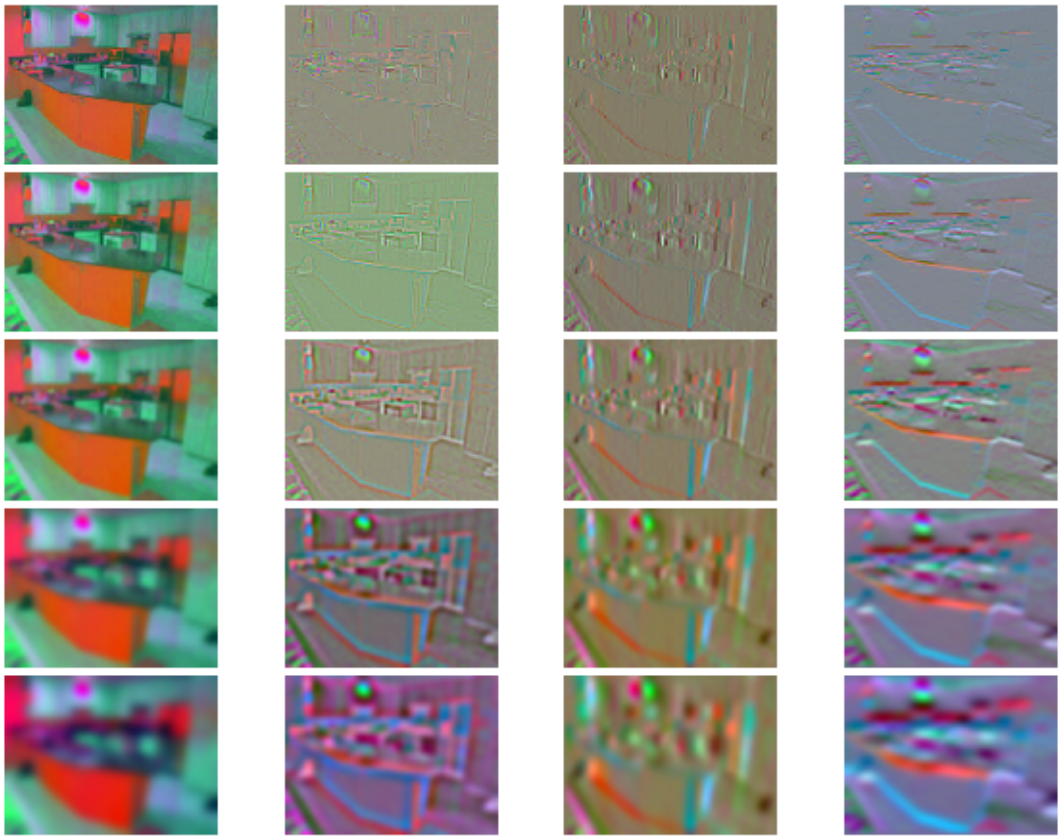

Figure 1: Image convolved with all 20 filters

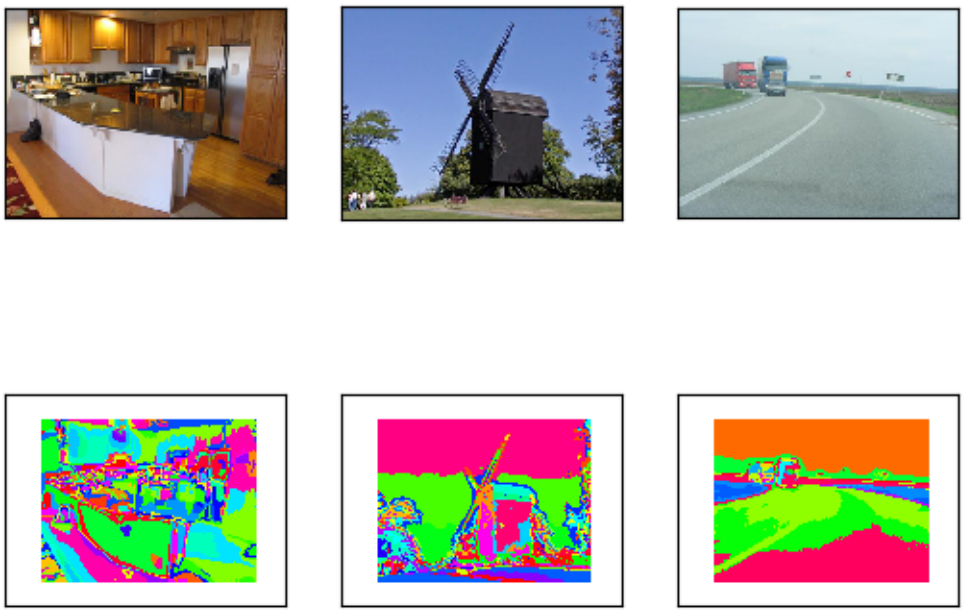

Figure 2: Images and corresponding wordmaps

2 Question 2

2.1 Q2.1

Code is included in the submission folder.

2.2 Q2.2

Code is included in the submission folder.

2.3 Q2.3

Code is included in the submission folder.

2.4 Q2.4

Code is included in the submission folder.

2.5 Q2.5

The confusion matrix is give below:

```
C = [[12  0  1  0  1  4  2  0]
 [ 1  9  1  2  0  1  0  6]
 [ 1  0 10  2  0  2  0  5]
 [ 0  1  2 13  0  0  1  3]
 [ 7  1  0  1  9  0  2  0]
 [ 3  0  0  0  5 12  0  0]
 [ 1  0  0  1  0  4 13  1]
 [ 1  1  1  6  0  0  0 11]]
```

The accuracy of the system is **55.625%**

2.6 Q2.6

Taking the folders in ascending order with 0 being auditorium and 7 being windmill, the most difficult classes for my system to classify where the following pairs [(actual_label, test_label)]

1. (0,4) – (auditorium, kitchen)
2. (1,7) – (baseball_field, windmill)
3. (7,3) – (windmill, highway)

There are multiple reasons for the pairs not having correct labels during test:

Smaller Number of Clusters: The first one is that the dictionary does not have sufficient words to correctly classify segments of the image, and hence attributes it to other words and may end up classifying wrongly - [Case 1 above]

Sampling: The random permutation for α pixels may end up in image patches that have zero or no coherent information for the classifier to compare against in the dictionary.

Semantic Similarity: Many images have segments that are similar like highways and windmills have large open spaces and the wordmap may end up really similar to each other - resulting in misclassification of the image ([Cases 2,3 above])

Some sample images for misclassification are given below:

Figure 3: Windmill that was misclassified as highway. There are many open spaces both at th top and the bottom of the image and the windmill has a perception depth to it which is similar to several highway images

Figure 4: Baseball_field that was misclassified as windmill. The image looks more similar to windmill settings where there is more grass and the there are structures in the middle of the image

3 Question 3

3.1 Q3.1

The code is included in the submission folder.

3.2 Q3.2

The confusion matrix is given below:

```
C = [[20  0  0  0  0  0  0  0]
 [ 1 17  1  0  0  0  1  0]
 [ 0  0 18  2  0  0  0  0]
 [ 0  0  0 20  0  0  0  0]
 [ 0  0  0  0 18  2  0  0]
 [ 0  0  0  0  2 18  0  0]
 [ 0  0  0  0  0  0 20  0]
 [ 0  0  1  1  0  0  0 18]]
```

The accuracy of the system is **93.125%**

Since the model is using pre-trained weights (from ImageNet dataset $\implies$ very large number of images for it to learn from), it has better understanding (more semantic information) of how images in scenes and objects in such scenes can look like and can build better representations internally for the image - making it easier for it to identify and classify the image.