16720A Computer Vision
Kartik Narula, knarula
Homework 1- Spatial Pyramid Matching for Scene Classification

Q1.1.1 (5 points): What properties do each of the filter functions pick up? (See Fig 3) Try to group the filters into broad categories (e.g. all the Gaussians). Why do we need multiple scales of filter responses? Answer in your write-up.

Ans. Each filter at a particular scale picks up different sets of features. The Gaussian filter picks up the signals of low frequencies and blocks the higher frequencies. The derivative of Gaussian in X direction picks the vertical edges. The derivative of Gaussian in Y direction picks the horizontal edges. The laplacian of gaussian corresponds to the second derivative and picks up the regions of rapid intensity change and edges.

Different scales pick up different kinds of features. Higher scales pick up broader features(eg. forests) while lower scales pick up narrower features(eg. trees). We need multiple scales to capture greater number of attributes from an image.

16720A Computer Vision
Kartik Narula, knarula
Homework 1- Spatial Pyramid Matching for Scene Classification

Q1.1.2 Apply all 4 filters at least 3 scales on aquarium/sun aztvjgubyrgvirup.jpg, and visualize the responses as an image collage as shown in Fig 4. The included helper function util.display filter responses (which expects a list of filter responses with those of the Lab channels grouped together with shape M * N * 3) can help you to create the collage. Submit the collage of images in your write-up.

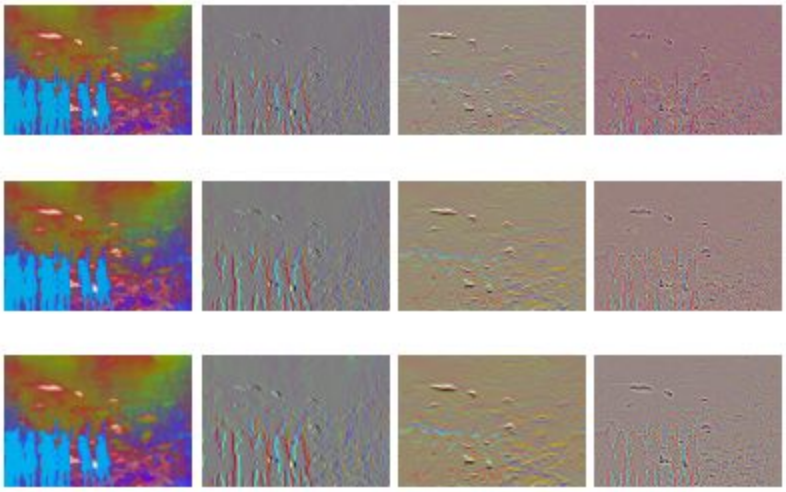
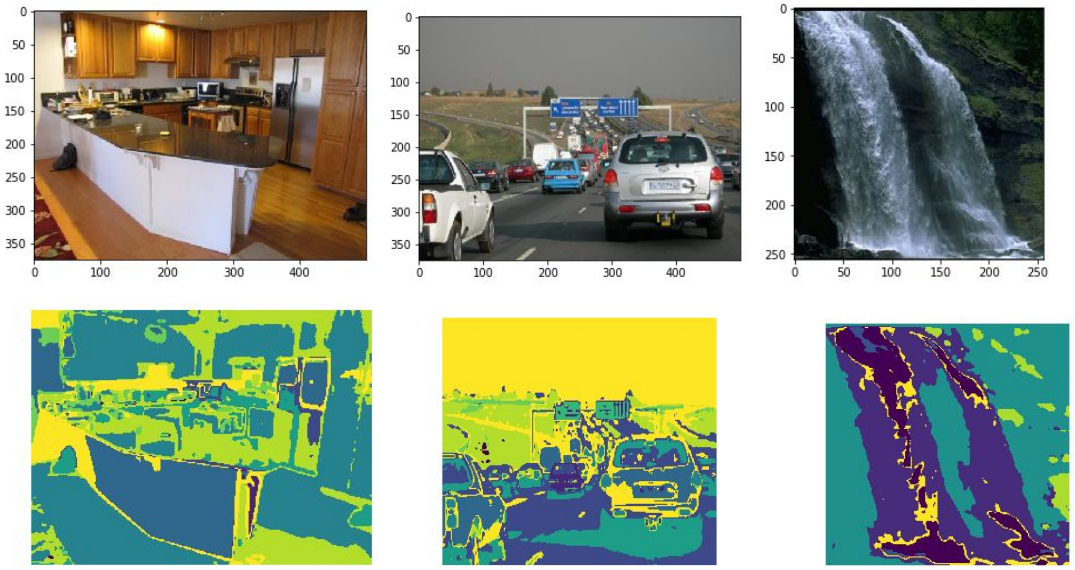


Fig1. Filter bank response on aquarium/sun aztvjgubyrgvirup.jpg

Each row represents one particular scale and each column represents a different type of filter. Gaussian filter(first column) blurs the image and omits the higher frequencies. Derivative of Gaussian in X direction(second column) picks up the vertical edges. The derivative of gaussian in Y direction(third column) picks up the horizontal edges. The laplacian of gaussian (fourth column) detects regions of rapid intensity changes.

16720A Computer Vision
Kartik Narula, knarula
Homework 1- Spatial Pyramid Matching for Scene Classification

Q 1.3. Visualize wordmaps for three images. Include these in your write-up, along with the original RGB images. Include some comments on these visualizations: do the \word" boundaries make sense to you?



I can differentiate different regions of the image by looking at the wordmap. I can identify the waterfall, car, kitchen shelf etc. from the wordmap. Similar pixels have been grouped together. Yes, the boundaries make sense.

16720A Computer Vision
Kartik Narula, knarula
Homework 1- Spatial Pyramid Matching for Scene Classification

Q2.5 (10 points):. Include the confusion matrix and your overall accuracy in your write-up. This does not have to be formatted prettily: if you are using LATEX, you can simply copy/paste it into a verbatim environment.

```
array([[23.,  3.,  5.,  1.,  2.,  3.,  7.,  6.],
       [ 1., 27.,  5.,  2.,  3.,  1.,  1., 10.],
       [ 2.,  1., 28.,  1.,  1.,  2.,  1., 14.],
       [ 4.,  2.,  3., 23., 11.,  1.,  5.,  1.],
       [ 1.,  3.,  4., 11., 19.,  7.,  3.,  2.],
       [ 3.,  0.,  4.,  1.,  2., 36.,  4.,  0.],
       [ 5.,  0.,  2.,  3.,  4., 10., 21.,  5.],
       [ 4.,  3., 10.,  1.,  3.,  4.,  3., 22.]])
```

Fig 2. Confusion Matrix with the default values

I achieved an accuracy of 49.75 % with the default set of parameters.

16720A Computer Vision
Kartik Narula, knarula
Homework 1- Spatial Pyramid Matching for Scene Classification

Q2.6 (5 points): In your writeup, list some of these hard classes/samples, and discuss why they are more difficult than the rest.

From the confusion matrix, images with the class-'park' have the highest levels of accuracy. This can be explained by the similar texture and color of the pixels in a park(trees, nature, sunlight etc).
Also, from the confusion matrix, the laundromat is classified as kitchen and vice versa a few times. This is because both types of images contain similar settings(tables, indoors), same lighting etc.

Windmills and highways are also classified as each other. This can be explained as both being outdoors, have pixels including sky etc.

16720A Computer Vision
Kartik Narula, knarula
Homework 1- Spatial Pyramid Matching for Scene Classification

Q3.1 (15 points): Tune the system you build to reach around 65% accuracy on the provided test set (data/test files.txt). A list of hyperparameters you should tune is provided below. They can all be found in opts.py. Include a table of ablation study containing at least 3 major steps (changing parameter X to Y achieves accuracy Z%). Also, describe why you think changing a particular parameter should increase or decrease the overall performance in the table you show.

- **filter scales: a list of filter scales used in extracting filter response;**
- **K: the number of visual words and also the size of the dictionary;**
- **alpha: the number of sampled pixels in each image when creating the dictionary;**
- **L: the number of spatial pyramid layers used in feature extraction.**

I achieved an accuracy of 49.75 % with the default values. I then increased alpha value to 35, which increased the accuracy to around 55%. I then increased the L value to 3, which lead to a drop in accuracy to 46%. I then increased the value of K to 15, and accuracy increased to 51 %. I kept playing with alpha and K values till I achieved 62.5% accuracy. Final values are as follows

Filters	K	alpha	L
[0.5,1,1.5,2,5]	25	350	3

- Increasing the value of alpha increases the accuracy because it increases the size of the training set. The model has more values to train.
- As the value of K increases, the model can classify pixels into more types.
- However, care should be taken, as very large values of K and alpha would lead to overfitting.
- I saw a decline in accuracy with increasing the value of L. This can be because some of the spatial information might be redundant and hence does not contribute to the accuracy of the model.

16720A Computer Vision
Kartik Narula, knarula
Homework 1- Spatial Pyramid Matching for Scene Classification