Q 1.1.1

Gaussian Filter:

Gaussian Filter is a type of non-uniform low pass filter (removes high-frequency components) which is mainly used to blur the image thus helps in removing noise and detail and might not preserve image brightness.

Laplacian of Gaussian:

The LoG filter generally depicts the area of Intensity change, thus helps to pick up edges from an image.

Derivative of Gaussian in the X direction:

Helps to obtain the edges with respect to X axis

Derivative of Gaussian in the Y direction:

Helps to obtain the edges with respect to Y axis

Broad Classification:

Separable Filters: Gaussian, X and Y derivatives of Gaussian

Non-Separable Filter: Laplacian of Gaussian. (Second Derivative of Gaussian)

Need of Multiple Scales:

Trying out multiple scales helps to decide between the optimal scaled value and to figure out the right balance between **better noise removal and least loss of detail** from an image which gives a better and a desirable output.

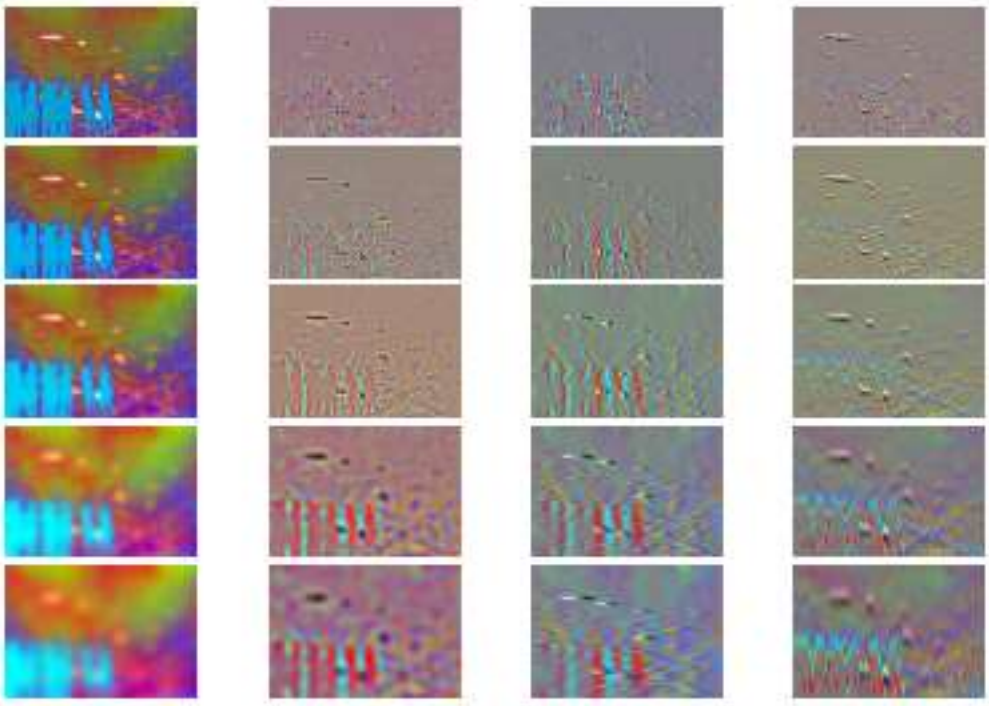
Q 1.1.2

Collage of filter responses:

The original image and the filter response of the image named **"sun_aztvjgubyrgvirup.jpg" shown** below.

Every row in the collage corresponds to the different scale of filters used and every column corresponds to the different types of filters.

The filters are in the order of **Gaussian, Laplacian of Gaussian, Derivative of Gaussian in the X direction, Derivative of Gaussian in the Y direction** respectively.



Q 1.2

Q 1.3 Visualisation of wordmaps:

Three images have been selected from the Kitchen dataset and their corresponding word Maps has been plotted.

Colour map used for plotting wordmap: '**Spectral'**






Yes, the word boundaries do make sense. Every word boundary corresponds to a feature in the image and it is differentiated by variety of colours. For E.g., in the image 2, the boundary around the person is created along the edges interpreting the feature correctly.

Q 2.1

Q 2.2

Q 2.3

Q 2.4

Q 2.5 - Spatial Pyramid Matching

Accuracy and Confusion Matrix:

```
[[ 9.  0.  1.  1.  0.  1.  1.  1.]
 [ 0. 15.  0.  1.  0.  2.  0.  0.]
 [ 1.  3.  5.  5.  1.  4.  2.  4.]
 [ 1.  2.  2. 14.  0.  3.  2.  2.]
 [ 0.  0.  0.  0. 11.  1.  1.  0.]
 [ 3.  1.  0.  2.  9.  8.  1.  0.]
 [ 1.  5.  0.  1.  3.  1. 10.  0.]
 [ 1.  1.  2.  3.  0.  1.  2.  9.]]
```

Accuracy:

0.50625

Q 2.6 Wrong classification - Hard examples :

1.

The Laundromat class was accidently classified as Kitchen several times .This was observed from the confusion matrix and the count was 8 meaning it was wrongly classfied 8 times.

One of the main reasons for the misclassification was the shape of the items in the images which were identical

The following image proves it.

This picture can be misjudged even my humans.

2.

Highway is wrongly classed as most of the other classes like Windmill and park.

I feel that the sky plays a major role in the misclassification of image as It is common in all the wrongly classfied images.

Q 3.1

Q 3.2 - Vgg16 Network : Confusion Matrix and Accuracy.

```
[[14.  0.  0.  0.  0.  0.  0.  0.]
 [ 0. 17.  0.  0.  0.  0.  0.  1.]
 [ 0.  0. 24.  0.  0.  0.  0.  1.]
 [ 0.  0.  0. 26.  0.  0.  0.  0.]
 [ 0.  0.  0.  0. 12.  1.  0.  0.]
 [ 0.  0.  0.  0.  1. 23.  0.  0.]
 [ 0.  0.  0.  0.  0.  0. 21.  0.]
 [ 0.  0.  0.  0.  0.  0.  0. 19.]]
```

Accuracy:

0.975

The results obtained are **better than** the classical BoW.

The primary reason is because in CNN, especially Vgg16, the network was trained using Millions of Images belong to the ImageNet dataset and has processed through hours of training and learning the Features of the images. Thus, the feature vectors obtained were close to perfection and thus the accuracy was twice the accuracy obtained through Bag of Words Approach. Also, there is no training or learning taking place in BoW unlike the neural network layers.