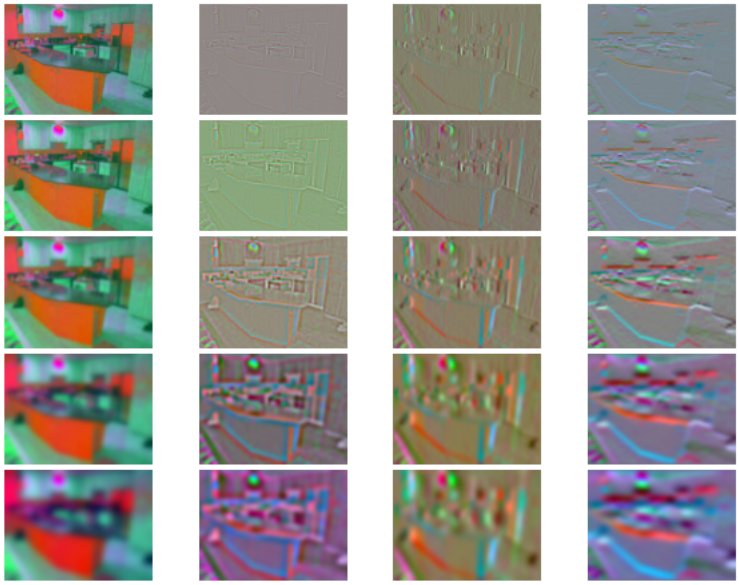
16-720 Computer Vision: Homework 1 (Fall 2022)
Spatial Pyramid Matching for Scene Classification

Haejoon Lee

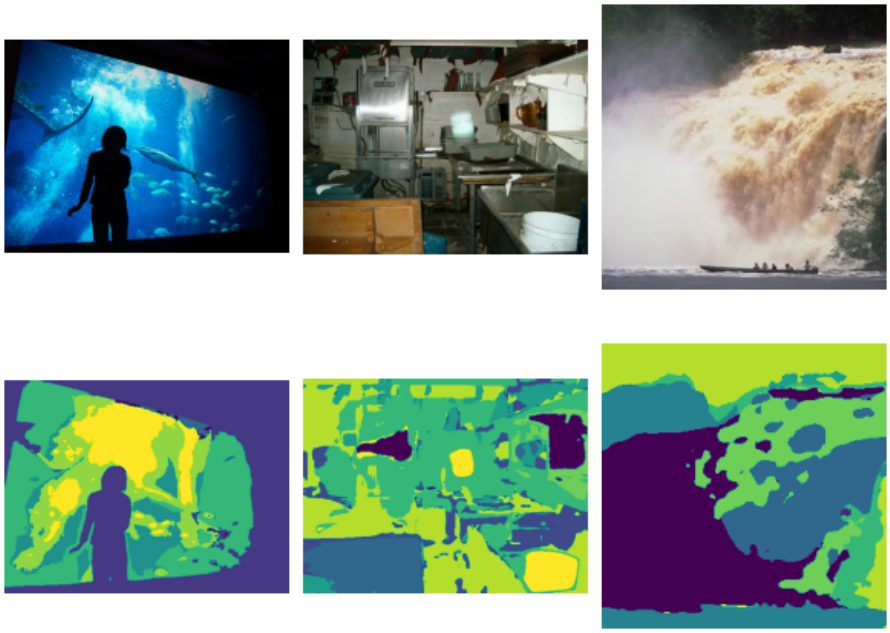
1.1.1

Gaussian Filter blurs images and removes noise. Laplacian of Gaussian filter is used for picking up rapid intensity change in images through 2nd derivative operation, which is useful for capturing edges. X derivative and y derivative of Gaussian filter also capture edges along each x and y direction. All the four filters could be driven from the Gaussian filter, but the Gaussian filter is for image smoothing and reducing noise, while the other three filters are for capturing edges. Since the size of edges in images could vary, we need multiscale filters to detect them efficiently.

1.1.2



1.3



2.5

```
In [1436]: conf
```

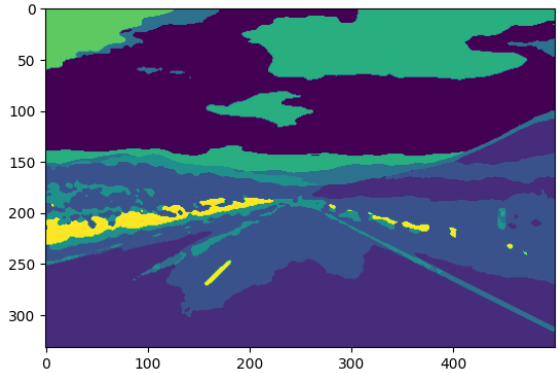
```
array([[24.,  3.,  5.,  5.,  5.,  0.,  3.,  5.],
       [ 1., 31.,  4.,  5.,  2.,  1.,  1.,  5.],
       [ 0.,  7., 19.,  2.,  4.,  4.,  2., 12.],
       [ 6.,  3.,  1., 19., 10.,  4.,  5.,  2.],
       [ 4.,  1.,  3.,  5., 31.,  4.,  1.,  1.],
       [ 3.,  0.,  2.,  3.,  3., 26.,  5.,  8.],
       [ 1.,  1.,  4.,  0.,  8.,  5., 27.,  4.],
       [ 4.,  6., 11.,  1.,  2.,  3.,  1., 22.]])
```

```
In [1435]: accuracy = np.trace(conf)/conf.sum()
           accuracy
```

```
0.4975
```

2.6

As shown in the above confusion matrix, 11 among 49 highway images (lable = 2) were classified to windmill (label = 7)



Suppose a windmill is small like the below image. In that case, the filter histogram of the image must be very close to highway images in the bags-of-words approach since major filter responses will represent the image by histograms.

3.1

filter_scales	K	Alpha	L	Accuracy
[1, 2, 4, 8, 8*sqrt(2)]	10	25	2	49.75%
[1, 2, 4, 8, 8*sqrt(2)]	10	50	2	51.75%
[1, 2, 4, 8, 8*sqrt(2)]	20	50	2	42.50%
[1, 2, 4, 8, 8*sqrt(2)]	10	1000	2	48.50%
[1, 2, 4, 8, 8*sqrt(2)]	10	100	2	48.25%
[1, 2, 4, 8, 8*math.sqrt(2), 16]	10	1000	2	46.75%
[1, 2, 4, 8, 8*sqrt(2)]	10	50	4	49.00%
[1, 2, 4, 8, 8*sqrt(2)]	200	100	3	57.50%
[1, 2, 4, 8, 8*sqrt(2)]	250	150	3	53.45
[1, 2, 4, 8, 8*sqrt(2)]	400	400	2	**59.25%**

Increasing alpha led to increase in accuracy. It is expected that increase of training data made the features represent the images better. Otherwise, increase of K from 10 to 50 decreased accuracy. It could be due to that 20 dictionaries might be too many to correctly represent each cluster in the feature space with alpha=50. However, when I increased K and alpha together (K = 400, alpha=400), the model provided the highest accuracy. I guess number of dictionaries could be matched with the number of training data. Increasing L higher than 2 didn't ensure performance improvement as reported in the original paper (S. Lazebnik, C. Schmid, and J. Ponce, Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories, CVPR 2006).

3.2

To make further improvements, I make all the input images to be the same size (512x512) by cropping and interpolation to ensure an equal number of pixels for each cell in the Spatial Pyramid Matching layers. I expected that this will generate more applicable histograms for each layer and make histogram comparison more accurate.

```
In [134]: conf2

array([[21.,  4.,  4.,  6.,  3.,  3.,  1.,  8.],
       [ 0., 33.,  5.,  4.,  3.,  0.,  1.,  4.],
       [ 1.,  6., 24.,  1.,  0.,  5.,  2., 11.],
       [ 3.,  4.,  2., 21., 12.,  4.,  3.,  1.],
       [ 2.,  2.,  2., 12., 23.,  5.,  1.,  3.],
       [ 3.,  0.,  5.,  1.,  5., 27.,  2.,  7.],
       [ 1.,  0.,  0.,  1.,  5.,  6., 30.,  7.],
       [ 3.,  3.,  9.,  1.,  5.,  4.,  4., 21.]])
```

```
In [136]: accuracy = np.trace(conf2)/conf2.sum()
          accuracy

0.5
```

Model	filter_scales	K	Alpha	L	Squalize	Accuracy
Vanila	[1, 2, 4, 8, 8*sqrt(2)]	10	25	2		49.75%
Vanila + Squalize	[1, 2, 4, 8, 8*sqrt(2)]	10	25	2		**50.00%**

As a result, the technique improved the total accuracy by 0.25%. It is noteworthy that it improved the model's accuracy, especially on hard samples (e.g., label=3, from 38.78% to 48.98%). It can be explained that the squalization reduce the importance of large scene such as the sky, which could not be critical for classification.