16720-B Computer Vision

NITHIN KUMAR SATHISH

DEPARTMENT OF ELECTRICAL ENGINEERING

nsathish@andrew.cmu.edu

Q1.1.1 Extracting Filter Responses

Gaussian filters are used to smoothen the image i.e each pixel's new value is weighted average of the pixels neighborhood. Typically used to reduce noise. It is low-pass filter removing high frequency components.

Laplace Gaussian are derivative filters used to find areas of rapid changes in images (edges). It is the second order derivative of the image. It's an isotropic filter i.e it applies equally in both x and y direction. The Laplacian is applied on an image which is smoothened by Gaussian in order to reduce its sensitivity to noise

Derivative Gaussian of gaussian in the x direction captures the areas of rapid changes along the x direction.

Derivative Gaussian of gaussian in the y direction captures the areas of rapid changes along the y direction.

We can broadly classify this into two groups. Gaussian Filters are used to smoothen the image it's not a derivative filter. The other three are derivative filters. Laplace Gaussian is a second derivative filter

We need multiple scales of filter responses because an image has lot of information and there could a lot of high frequency and low frequency edges. So in order to capture both edges in the image we use different scales.

Q1.1.2 Visualization of Filter Responses

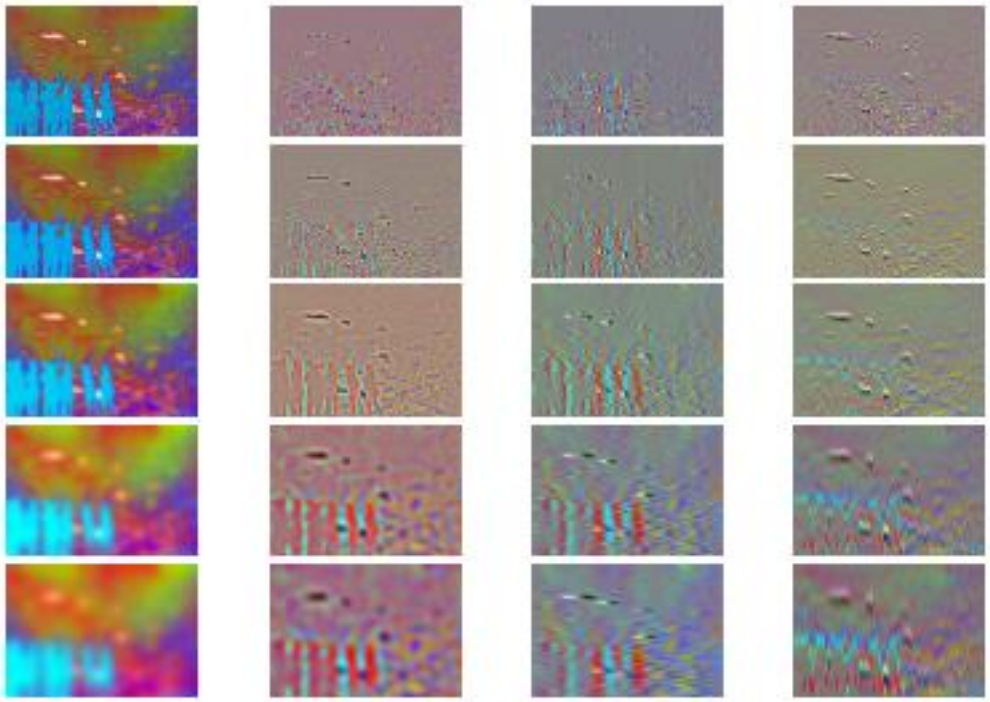
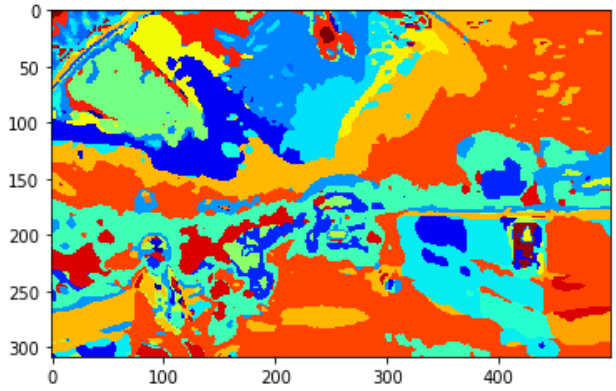
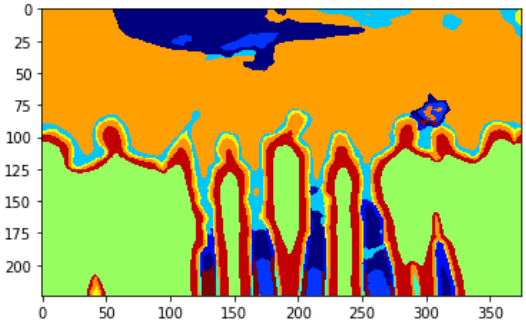
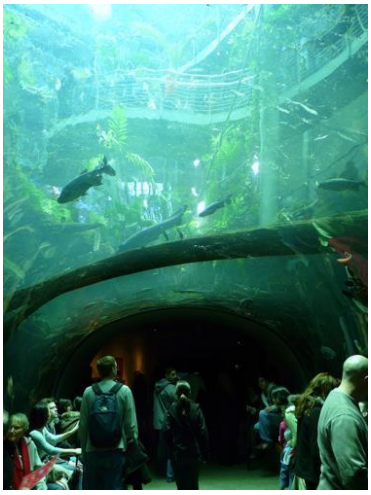


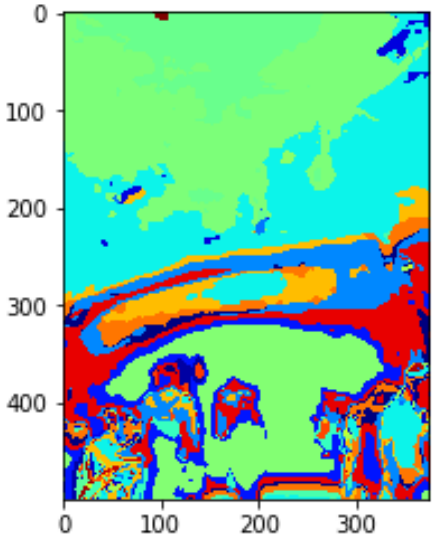
Fig1. Visualization of Filter banks on aquarium/sun_aztvjgubyrgvirup

Q1.3 Visualization of Wordmap









We can see that each pixel in the wordmap is assigned to its closest word in the dictionary learned. The dictionary in this case has learnt broad textures in the original image like humans, fishes etc.

Q2.5 Confusion Matrix and Accuracy Value

Accuracy=50.625

Confusion matrix

[[9. 0. 0. 2. 1. 1. 0. 1.]

[0. 12. 0. 2. 0. 1. 3. 0.]

[1. 1. 12. 1. 2. 2. 1. 5.]

[0. 6. 1. 9. 1. 5. 1. 3.]

[3. 1. 0. 0. 9. 0. 0. 0.]

[1. 5. 2. 0. 7. 8. 1. 0.]

[1. 0. 0. 0. 0. 4. 15. 1.]

[1. 1. 1. 5. 1. 1. 2. 7.]]

Q2.6 Failures in Bag of Words

Predicted:Kitchen

Actual:Desert

Predicted: Kitchen

Actual: Laundromat

Predicted: Aquarium

Actual: Waterfalls

It can be seen from the above examples that the algorithm depends a lot on the colors, edges and textures. Since the kitchen and the laundromat have similar color and texture it gets misclassified.

The desert doesn't does not have enough edges and since the color is similar to kitchen it gets misclassified as kitchen

Similarly, the waterfalls has similar texture and color as the aquarium so that gets misclassified too.

Q3.2 Deep Learning Features

Accuracy= 97.5

Confusion matrix=

[[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.]]

The results obtained from the Deep Learning is better compared to the Bag of Words approach. This is because by using the VGG architecture which has been pretrained on the Imagenet dataset, the features learnt in all the layers combine to give more optimal features. In the Bag of Words approach the filters which we used are generic and is not able to capture the most important features necessary to distinguish between scenes. It mostly tries to distinguish between images through colors, textures and edges which is not optimal in this dataset.