

Semantic Scene Segmentation Using Bag Of Words

2018-22635 임형석

1. Problem Definition

본 과제에서는 Bag of words를 이용하여 scene segmentation을 얻는 것이 궁극적인 목표이다. 이러한 scene segmentation을 얻기 위해 여러가지 단계를 거치는데 일단 이미지의 feature를 뽑기 위하여 Gaussian Kernel, DoG, LoG등 총 17가지의 filter를 이용해서 얻은 texton feature를 구하여 kmeans 알고리즘을 통해 superpixel을 histogram으로 나타내준다. 이렇게 superpixel을 histogram으로 나타내주고 그것을 superpixel label histogram과 함께 Markov Random field Energy를 얻은 후, graph cut으로 MRF Energy function을 최적화해주면서 Scene segmentation을 해주는 프로젝트이다.

데이터는 MSRC v1 dataset을 사용하여 총 14개의 object로 이루어져 있고, train 이미지 130개 test 이미지 110개로 총 240개의 이미지와 ground truth 240개의 이미지가 존재한다.

Texton feature를 얻기 위하여 총 17가지의 filter를 bank로 생성하였는데, Gaussian filter ($\sigma = 1,2,4$), Laplacian of Gaussian filter ($\sigma = 1,2,4,8$), Differential of Gaussian filter ($\sigma = 2,4$)(horizontal, vertical) 와 같은 filter들을 생성하여 전체 image에 filter bank를 convolution하여 texton feature들을 얻었다.

위에서 얻은 texton feature 중 train image에 해당하는 feature를 kmeans clustering 알고리즘에 데이터로 사용하여 각 픽셀을 100가지의 class로 clustering 할 수 있게 kmeans 모델을 생성해주었다.

그리고 각 이미지의 superpixel을 얻기 위하여 slic 라이브러리를 이용하였고 Bag Of Words를 사용하여 각 superpixel을 100차원의 histogram으로 표현할 수 있게 생성해주었다. 그렇게 생성한 histogram들과 label을 바탕으로 superpixel의 MRF Energy function을 다음 식과 같이 구하였다. $E(l) = \sum_p f(l_p) + \sum_{(p,q) \in N} \theta_{pq}(l_p, l_q)$

마지막으로 위에서 구한 energy function을 바탕으로 graph cut을 이용해 최적화를 시켜주면서 image가 object 개수만큼 잘 segment 되도록 하였다.

2. Dependencies

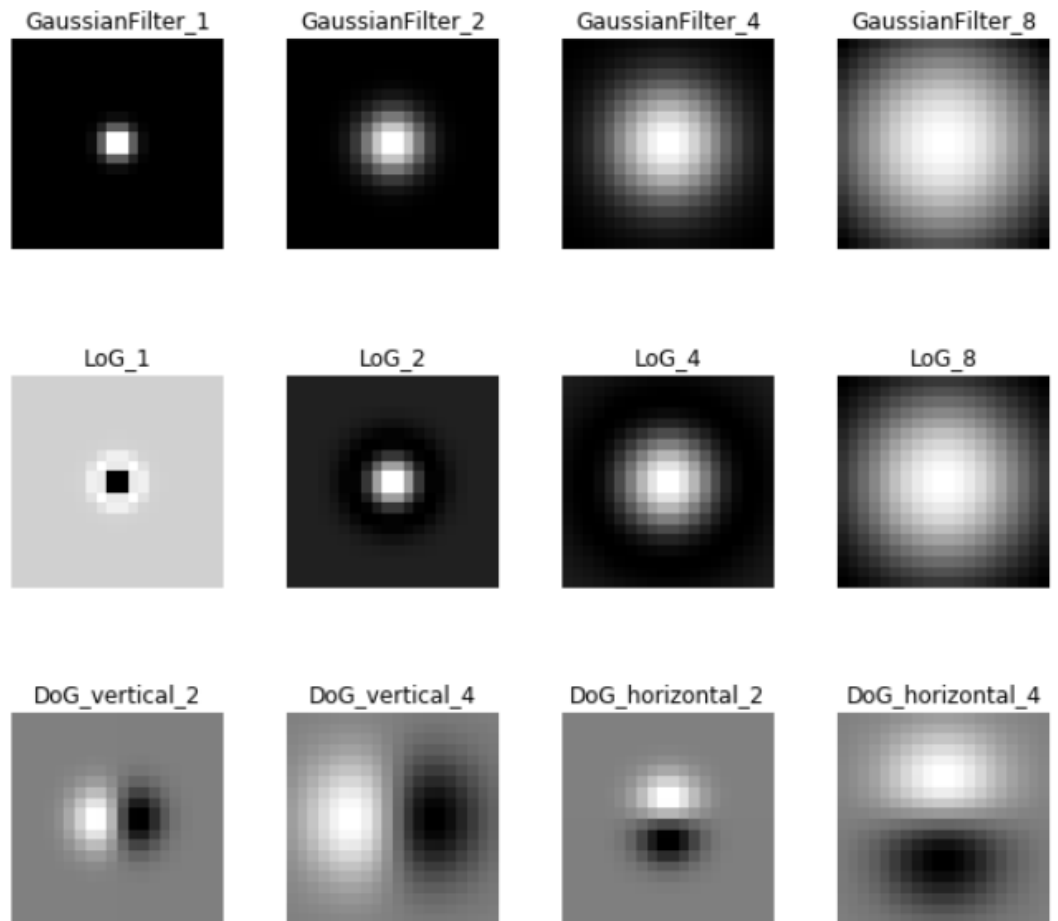
- A. Ubuntu 18.04
- B. Python 3.6.8

- C. Numpy==1.16.2
- D. Matplotlib==3.0.3
- E. Cv2==4.0.0
- F. Opencv-python==3.4.6
- G. Opencv-contrib-python==3.4.6
- H. Scipy==1.2.1
- I. Sklearn==0.0
- J. Scikit-learn==0.20.3
- K. Scikit-image==0.14.2
- L. Pickleshare==0.7.5

3. Results

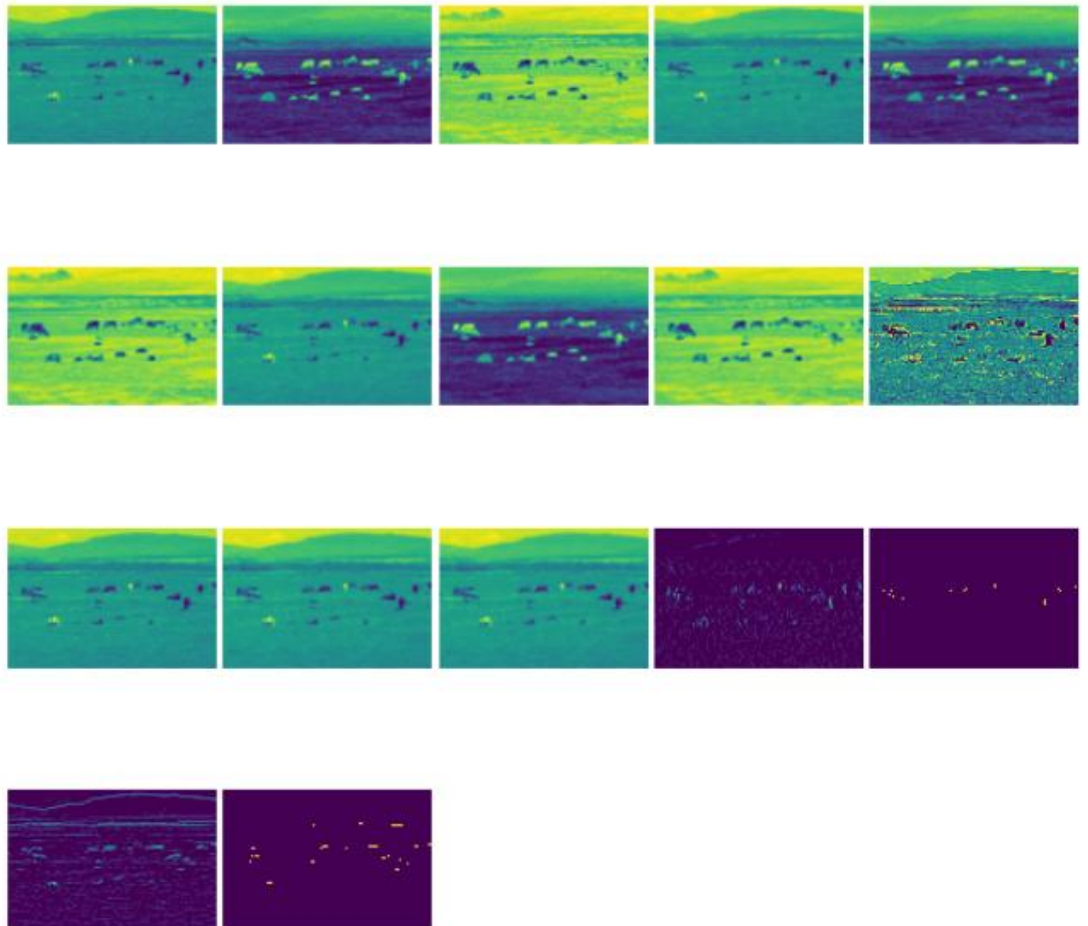
- A. Filter banks for texton filter

Filter shape 20x20 (for visualize)



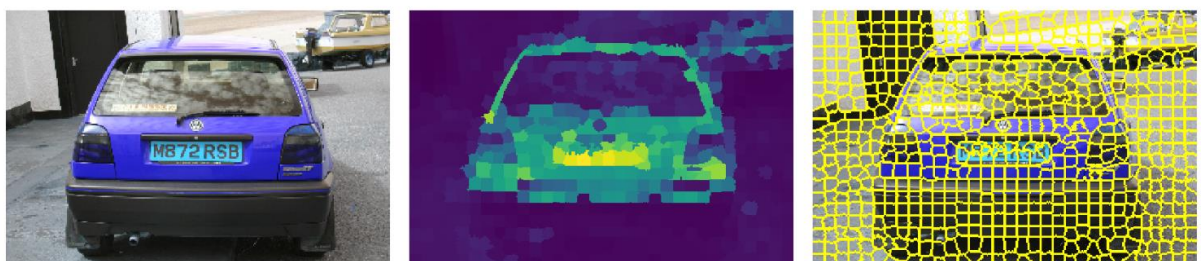
B. Texton feature for train image

Textron feature example



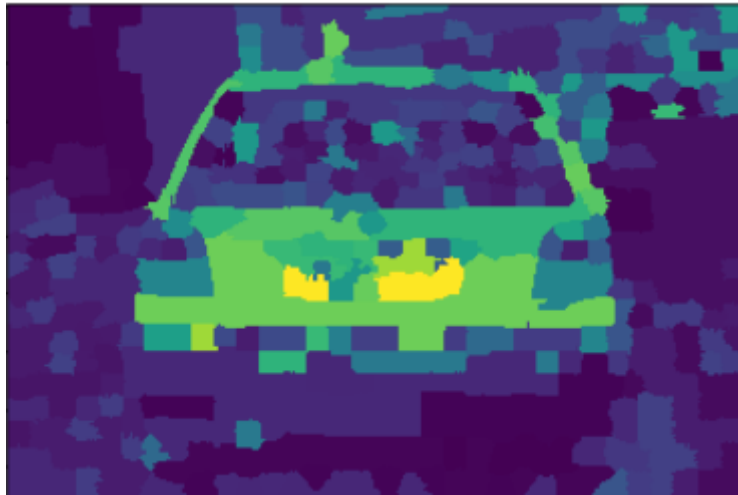
C. Superpixel Representation

Superpixel



D. BoW Representation

- i. Restored Image by BoW representation (Histograms)



E. Result of Scene Segmentation

- 미구현

F. Per-pixel classification accuracy

- 미구현

G. Per-class classification accuracy

- 미구현