

Assignment #4 report

2022311113 이소현

I. Intro

UKBench dataset의 2000개의 이미지에서 추출된 SIFT features와 CNN features를 이용하여 image descriptor를 생성하는 코드를 만들었다. 각 descriptor는 4096차원을 가지고 있다.

II. Method

1. K-Means clustering

SIFT features와 CNN features에 대하여 각각 K-Means clustering을 적용하였다. SIFT feature의 경우 128차원으로 이루어져 있기 때문에 cluster의 개수를 32 ($4096/128$)로 설정한 다음, 20만 개의 feature를 랜덤으로 추출하여 32개의 centroid를 구했다. 이후 모든 SIFT feature에 대하여 각 centroid와의 거리를 구하여 가장 가까운 centroid가 속해 있는 cluster에 할당하였다. CNN feature의 경우 각 이미지마다 (14, 14, 512) 형태로 되어있었기 때문에 (196, 512) 형태로 resize하는 과정을 추가하였다. Resize한 후의 feature는 512차원이므로 cluster의 개수를 8 ($4096/512$)로 설정한 다음 8만개의 feature를 랜덤으로 추출하여 8개의 centroid를 구한 뒤 전체 CNN feature들을 알맞은 cluster에 할당하였다.

2. VLAD

K-Means clustering을 통해 구한 cluster를 이용하여 VLAD 벡터를 구하였다. 각 centroid와 매칭되는 features 사이의 누적 잔차를 계산한 뒤 flatten 과정을 통해 $k \times D$ size의 벡터를 만든 다음, 이 벡터들을 하나의 배열에 담기게 하고 L2 정규화를 적용하였다.

위의 방식으로 SIFT features만을 이용하여 생성된 배열과 CNN features만을 이용하여 생성된 배열의 평균을 구해 최종적인 image descriptors를 생성하였다.

III. Result

SIFT features만을 이용하여 구한 descriptors의 성능은 다음과 같이 나왔다.

```
PS C:\Users\이소현\Desktop\3학년\컴퓨터비전개론\CV_A4_2022311113>
A4_2022311113_SIFT.des 3.0705 (L1: 3.0705 / L2: 3.0655)
PS C:\Users\이소현\Desktop\3학년\컴퓨터비전개론\CV_A4_2022311113>
```

CNN features만을 이용하여 구한 descriptors의 성능은 다음과 같이 나왔다.

```
PS C:\Users\이소현\Desktop\3학년\컴퓨터비전개론\CV_A4_2022311113>
A4_2022311113_CNN.des 3.3340 (L1: 2.8770 / L2: 3.3340)
PS C:\Users\이소현\Desktop\3학년\컴퓨터비전개론\CV_A4_2022311113>
```

둘의 평균으로 최종적으로 구한 descriptors의 성능은 다음과 같이 나왔다.

```
PS C:\Users\이소현\Desktop\3학년\컴퓨터비전개론\CV_A4_2022311113>
A4_2022311113.des 3.6040 (L1: 3.5595 / L2: 3.6040)
PS C:\Users\이소현\Desktop\3학년\컴퓨터비전개론\CV_A4_2022311113>
```