HW1 Report

themis12@kaist.ac.kr / 010 2830 3651
20184448 Jo MinKi

1 Introduction

This work conduct the image searching task with neural network feature.

2 Result

Correct case

Query

dist = [3.6398244, 90.92494, 107.60222, 110.9077]

Error cases

Query

dist = [4.801503, 88.44673, 96.51866, 97.753426]

VGG19_fc6 layer

Summary
Total number of images: 1000
Number of correct entries: 943
Accuracy: 0.943

Time: 5.848515033721924sec

VGG19_fc7 layer

Summary
Total number of images: 1000
Number of correct entries: 940
Accuracy: 0.94

Time: 2.968522071838379sec

VGG16_fc6 layer

Summary
Total number of images: 1000
Number of correct entries: 927
Accuracy: 0.927

Time: 5.285831689834595sec

VGG16_fc7 layer

Summary
Total number of images: 1000
Number of correct entries: 943
Accuracy: 0.943

Time: 2.549124240875244sec

VGG16_pool5 layer

Summary
Total number of images: 1000
Number of correct entries: 399
Accuracy: 0.399

ResNet v1_50 block5 conv3 layer

Summary
Total number of images: 1000
Number of correct entries: 330
Accuracy: 0.33

According to the observation, feature FC6 layer of the VGG model dramatically works better than the feature from the end of the ResNet, VGG pooling layer (or the end of the conv layer). In addition, the processing time took around 5.5 second for 250 queries when you use the L2 distance. If you use L1 distance, it will takes around 2.5 sec for 250 queries.

VGG model has the Fully Connected layer which is avoided by recent trend because FC layer prone to make the model overfit. However, if the model does not have the FC layer at all, such like a ResNet or Inception net or VGG Net bellow the Pool5 layer, there is no pixel-wise non-linear operation. In other words, the fc layer makes the model be robust on the rotation. However, the strength of the ResNet is the capacity of the expressional dimension of the function. Which means that the ResNet is able to express more difficult function, so if you train the model in the searching purpose, it can get better.

3 Discussion

a. Approach

The problem of this is that this task used the Neural network model but this model was not trained for the searching task. In other word, there was no Training.

If there are enough data for this task, the model performance can be improved by adding one more embedding layer or replacing the FC layer for the image embedding purpose. Image embedding or Metric learning concepts can be applied for the improvement.

b. Exhaustive search

For this assignment, any compression or coding method and fancy data structure has been used. It was fine for this assignment which require searching only 1000 images. However, when the number of the image gets bigger such like a search engine, this method will not work.

There could be a solution such like quantization. You can quantize the weight and find the cluster based on the distance. If you choose the exhaustive way for quantization, then it also takes n^2 times, but it must be much faster when you conduct a search. (If you group each portion with 256 images, it will only take 256 time. Furthermore, there are many ways that conduct the quantization in $n \log n$ time.