CV HW5 Report

Group 13

Introduction

- Try different feature representations and classifiers to categorize images into one of 15 scene types
- Use two feature representations: tiny images representation and bag of word model
- Use two different classifier: KNN and SVM

Tiny images representation

Tiny images representation make image be a 16*16 feature vectors and it has two categories:

- Resize

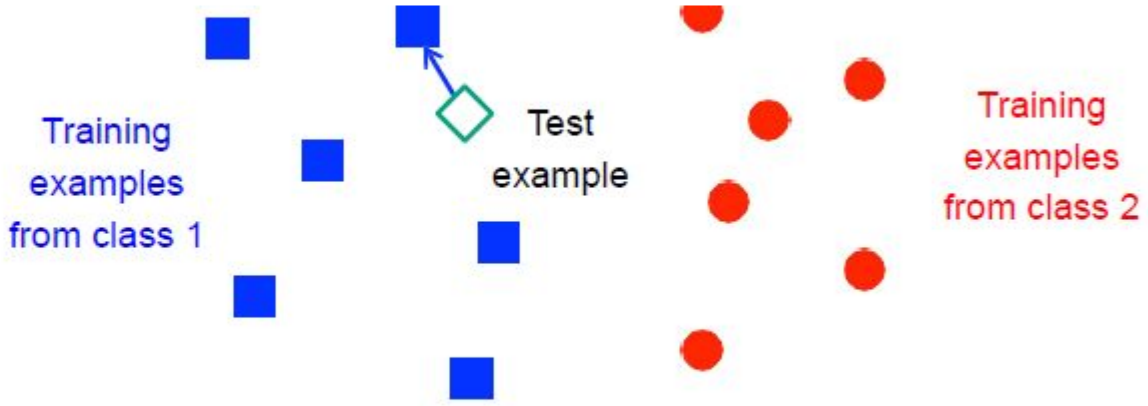
Resize image to a 16*16 fiexed resolution

- Crop

Crop the 16*16 regions from the center of image

KNN

- Find the k nearest datas. In our work, we use the SSD to find the nearest data
- Let k nearest datas vote which class has most number



Tiny images representation + KNN

When load the image datas, we normalize the datas to make our result more accurate

In training stage, we use the tiny images representation to represent all the training image datas.

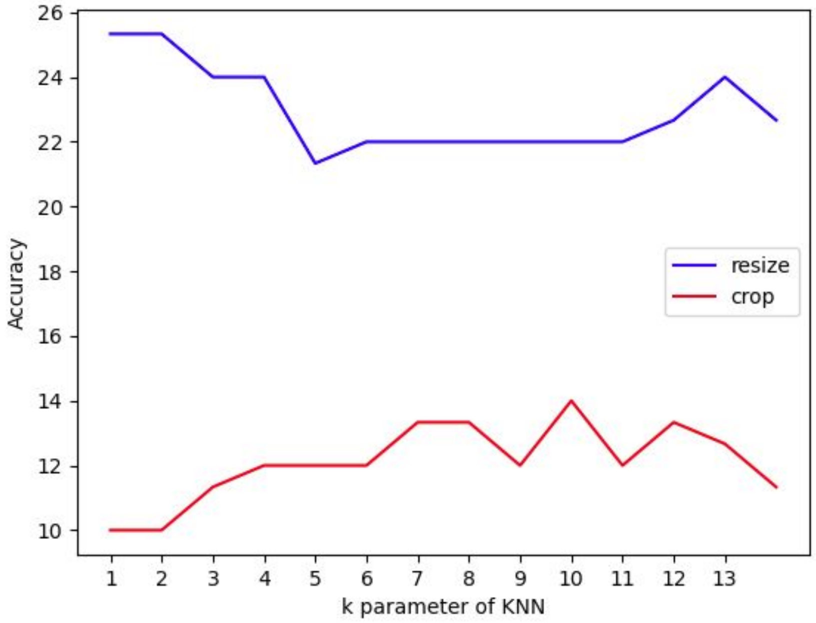
In testing stage, we also use the tiny image representation to represent the testing image. And using knn classifier checks which class belongs to the testing image

Experiment

In the experiment, we try the different numbers of k in KNN

From the result, we find that the accuracy of resizing image is better than croping image

The best accuracy is 25.33%



Bag of SIFT representaiton

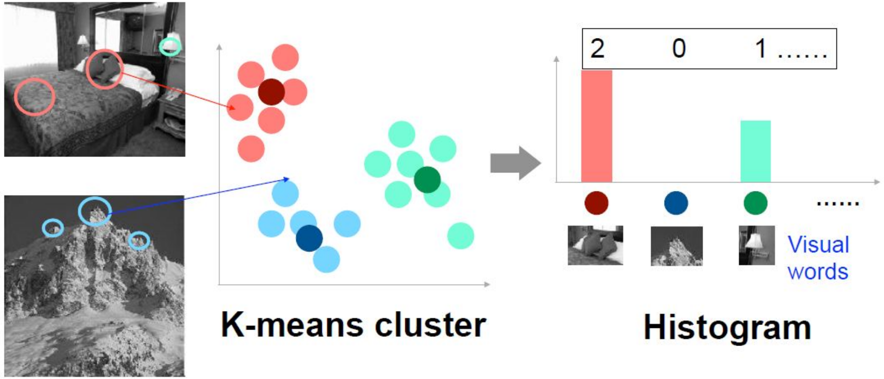
This feature representation uses SIFT feature and bag of word model. When loading the image datas, we resize the image to the 256*256. After resizing the images, We will form visual words by sampling many local features SIFT from our training set and then clustering them with kmeans

The number of kmeans clusters is the size of our vocabulary and the size of our features

Bag of SIFT representaiton

The traditional SIFT is sparse. In our work, we use dense SIFT. Why do we use the dense SIFT ? The dense SIFT is more suitable for image classification. We will explain more detial in the discussion

After generating SIFT descriptors and cluster centers, now we are ready to represent our training and testing images as histograms of visual words and the histograms are the bag of SIFT representations



Bag of SIFT representaiton + KNN

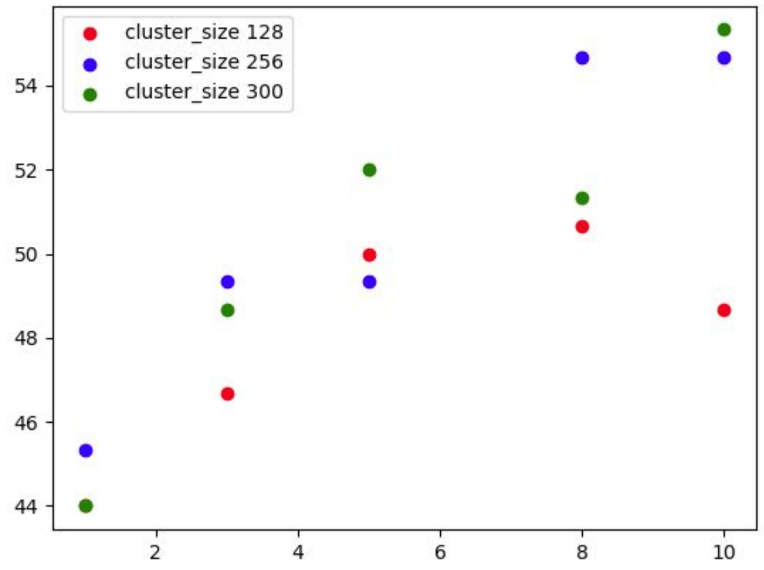
It is similar to tiny image representation + KNN. Use the bag of SIFT representation to represent the all image datas. And using knn classifier checks which class belongs to the testing image

Experiment

The clustering size 300 and the k 10 have the best result in our experiment

The step parameter of dense SIFT is 11

The best accuracy is 55.33%



Bag of SIFT representaiton + Linear SVM

This task is to train 1-vs-all linear SVMs to operate in the bag of SIFT feature space. Linear classifiers are one of the simplest possible learning models. The feature space is partitioned by a learned hyperplane and test cases are categorized based on which side of that hyperplane they fall on

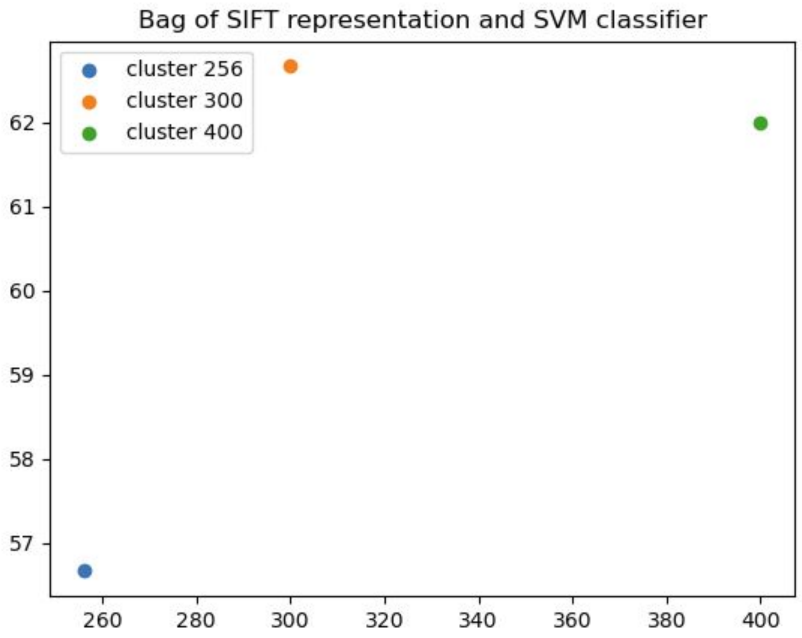
Despite this model being far less expressive than the nearest neighbor classifier, it will often perform better

Experiment

Try the different clustering size and tune the terminate threshold and cost of SVM to 5e-3 and 700.

The speed of SVM and the performance of SVM are all better than KNN

The best accuracy is 62%



Bonus - Deep Learning

Deep learning models can perform automatic feature extraction from raw data, also called feature learning.

In this task, we use the simple Resnet-18 model to perform the multi-classification tasks and implement it with pytorch

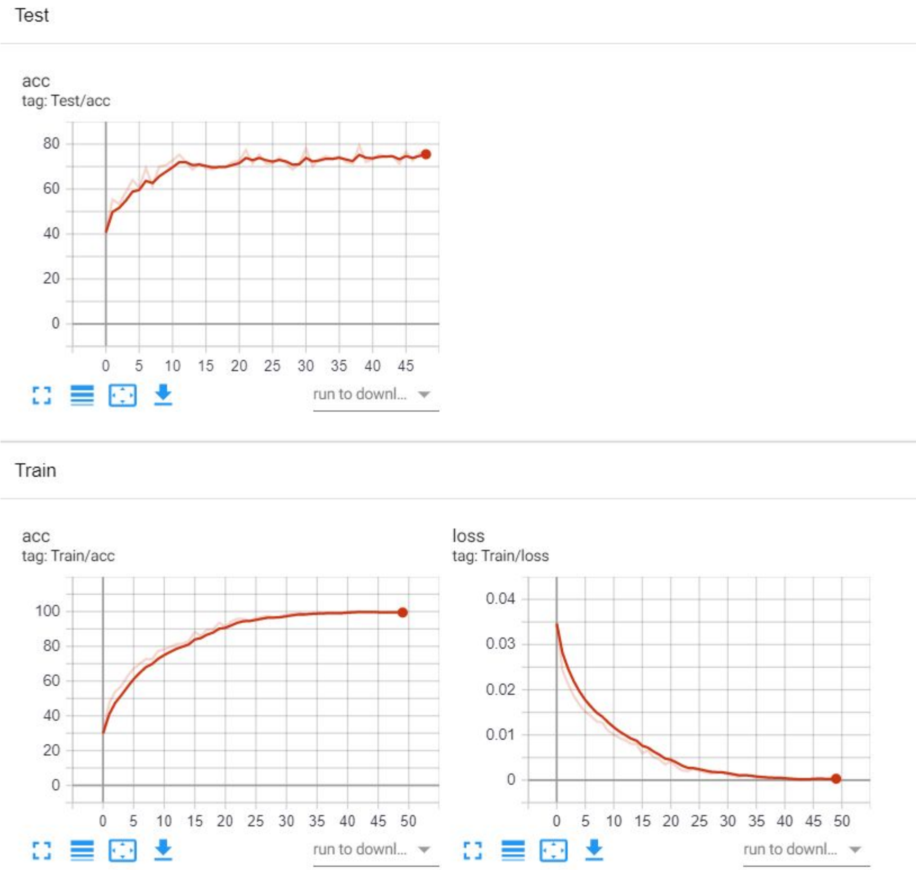
Experiment settings:

- lr: 0.001
- batch size: 64
- optimizer: Adam

Experiment

The best accuracy is 79.33%

We can see it outperform the other task that use the original hand-design features



Experiment

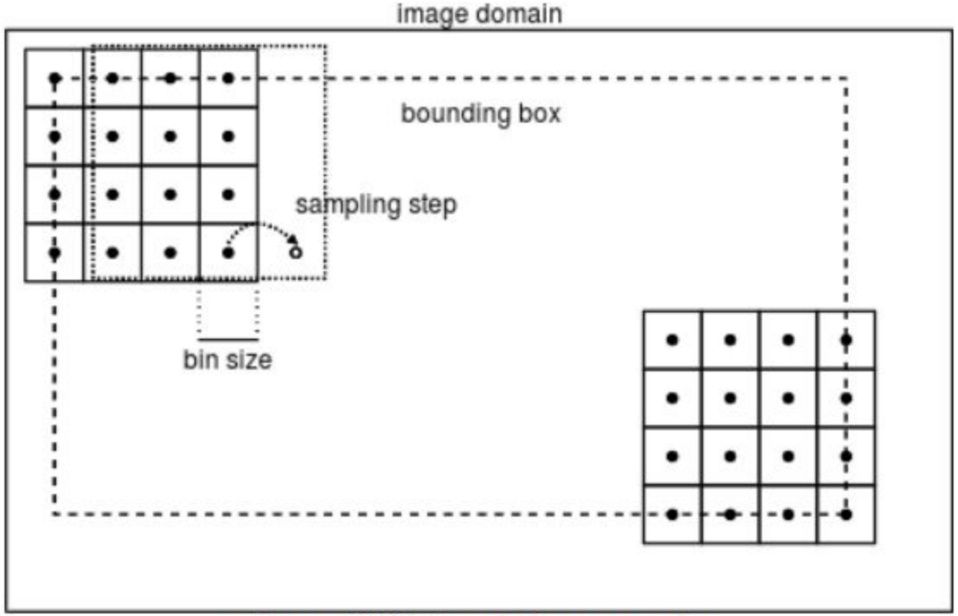
The evaluation of all tasks

tiny image + KNN	bag of SIFT+KNN	bog of SIFT+SVM	deep learning
25.33%	55.33%	62%	79.33%

Discussion - dense SIFT vs sparse SIFT

- Dense SIFT

Dense SIFT is roughly equivalent to running SIFT on a dense gird of locations at a fixed scale and orientation. This type of feature descriptors is often uses for object categorization



Dense SIFT descriptor geometry

Discussion - dense SIFT vs sparse SIFT

- Sparse SIFT

Spare SIFT can't represent the good feature representations in different object categories, so it isn't suitable for object categorization

But in the aspect of image searching, sparse SIFT is better than dense SIFT because the dense features can't be matching accurately

Discussion - cluster size

In the bag of SIFT representation, how to choose the clustering size is hard.

- Too small: the visual words are not representative enough
- Too big: the kmeans quantization is difficult and overfitting problem

In our experiments, we find the suitable clustering size is about 200~400. If the size is smaller than about 100, the performance will not be good

Conclusion

- Implement different feature representations, classifiers and measure their performance
- Dense SIFT is better than traditional SIFT in object categorization
- Deep learning is more powerful in image classification. It can learn better features automatically