Computer Vision Assignment Report

Title: Scene Recognition with Bag-of-Words

Student Name: 加泽瑄

Student ID: 12011126

1. Experimental Design

a. Tiny Image features with Nearest Neighbor classifier

To build a tiny image feature, simply resize the original image to a very small square resolution. One simple way is resize the images to square (16x16) while ignoring their aspect ratio.

For K-NN algorithm part, we choose k = 10 (less than experience value k = 13) to get a not-bad stable result. To get more possible better confusion, we randomly choose the possible points in range 10 of each point.

b. Bag of SIFT features with Nearest Neighbor classifier

The algorithm simply computes the distribution (histogram) of visual words found in the query image and compares this distribution to those found in the training images. The biggest difference from instance recognition is the absence of a geometric verification stage , since individual instances of generic visual categories, have relatively little spatial coherence to their features.

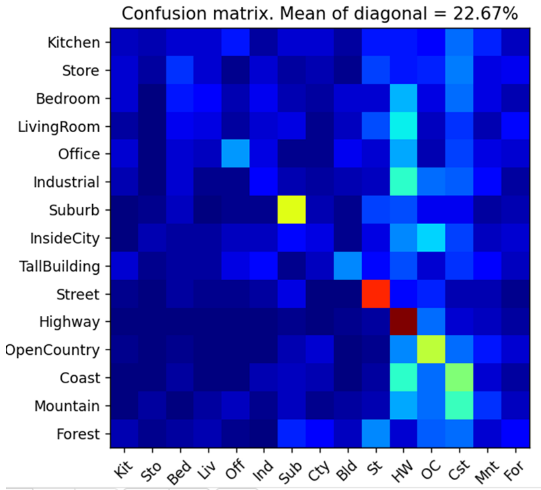
To improve the algorithm, a set of suitable parameter is very necessary. We designed a series of experiments where the step size of the dictionary and the step size of the model were adjusted each time during SIFT. We found that higher confidence can be obtained when the dictionary step size is larger than the model step size. However, if the model step size is too small, it will consume a lot of resources and time during calculation. Finally, we set the model step size to 5 and the dictionary step size to 9, which can reduce resource consumption as much as possible while breaking through the average.

c. Bag of SIFT features and SVM classifier

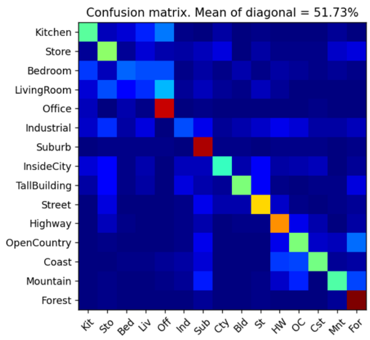
This part will train a linear SVM for every category (i.e. one vs all) and then use the learned linear classifiers to predict the category of every test image. Every test feature will be evaluated with all 15 SVMs and the most confident SVM will "win". Confidence, or distance from the margin, is W*X + B where* " is the inner product or dot product and W and B are the learned hyperplane parameters.

2. Experimental Results Analysis

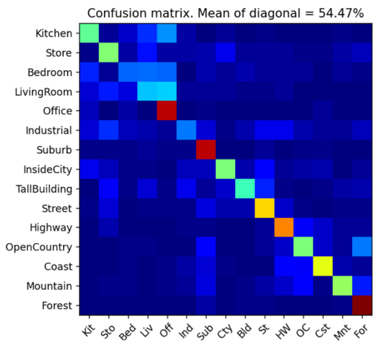
a. Tiny Image features with Nearest Neighbor classifier



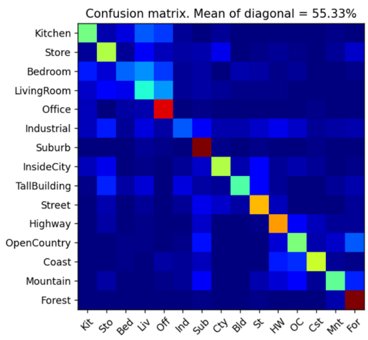
b. Bag of SIFT features with Nearest Neighbor classifier



Dic_step = 5
Get_step = 9

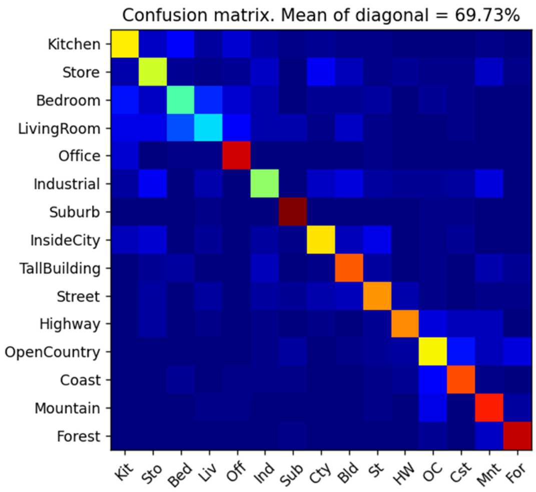


Dic_step = 7
Get_step = 7



Dic_step = 9
Get_step = 5

c. Bag of SIFT features and SVM classifier



3. Bonus Report (If you have done any bonus problem, state them here)

To improve the algorithm, a set of suitable parameter is very necessary. We designed a series of experiments where the step size of the dictionary and the step size of the model were adjusted each time during SIFT. We found that higher confidence can be obtained when the dictionary step size is larger than the model step size. However, if the model step size is too small, it will consume a lot of resources and time during calculation. Finally, we set the model step size to 5 and the dictionary step size to 9, which can reduce resource consumption as much as possible while breaking through the average.