University of British Columbia

CPSC 425: Computer Vision

Assignment 5

Simon Ghyselincks

Self-Studied based off of UBC CPSC 425 2023T1 course material

2023

Based off of UBC CPSC 425 2023W T1 course material found at `https://mattabrown.github.io/425/`

The goal is to explore image classification using some of the cutting edge techniques from 15+ years ago, namely bag of words and SVM. The images have a set of feature points and 128D SIFT descriptors. The descriptors are clustered using k-means to form a dictionary of words. These are basically textons that havd similarity and can be used to classify images.

The test images can be classified by applying sorting each of the descriptors into the matching k-means cluster from the training set.

To give a single descriptor vector to each image, we can use a histogram of the number of descriptors that match each cluster. This is a bag of words representation, where each cluster is a word and the image is a bag of words that has removed spatial information when it comes to the location of the feature points.

We inspect the average histogram for each of the 15 categories of images when it comes to the training set to get an idea of which images have a similar signature with this method. For this task only 60 cluster centers were used because it is very hard to visualize with a higher word count.
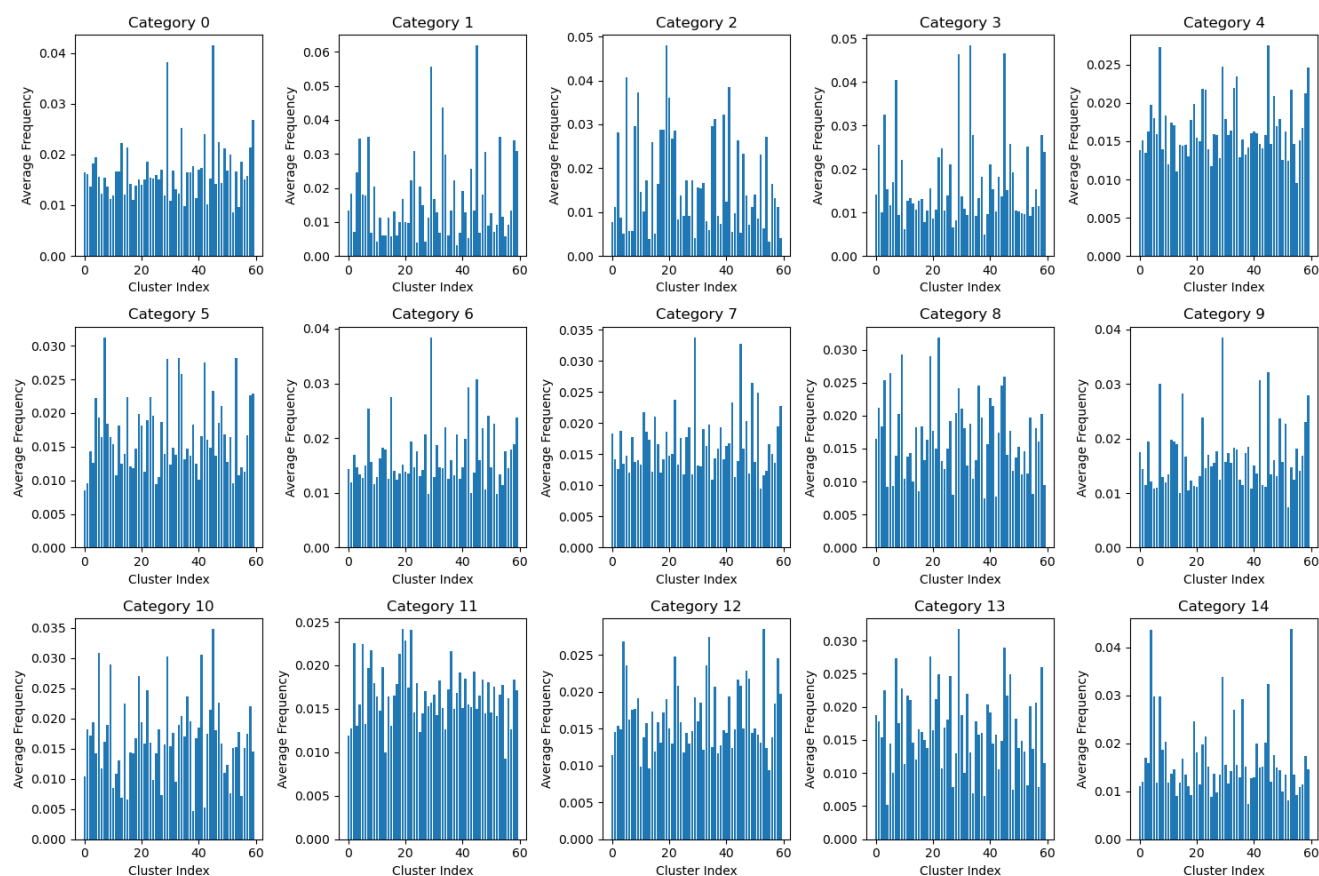

12-5 look similar and 12-4 as well. This is bassed on an overall shaper of the histogram.

Choosing K for KNN

Too low of a value at $k = 1$ overfits to the training data giving around 35% accuracy. Too high of a k when it reaches around 30 will also start to diminish the accuracy because of too much generalization. A good value is around $k = 15$, giving close to 40% accuracy.

The confusion matrix is as follows:

Confusion Matrix for KNN Classifier
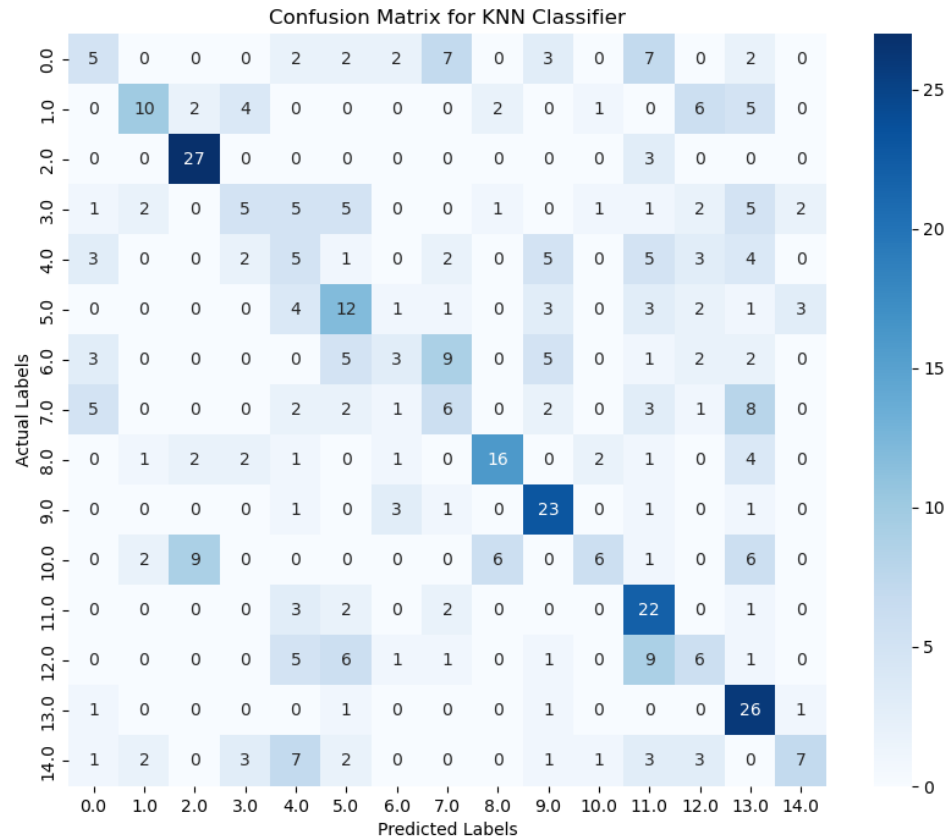
Choosing Regularization for SVM

Choosing a $C = 25$ improves the performance to around 49% accuracy. The improvement is due to the regularization of the SVM model which helps to prevent overfitting to the training data.

Confusion Matrix for SVM Classifier
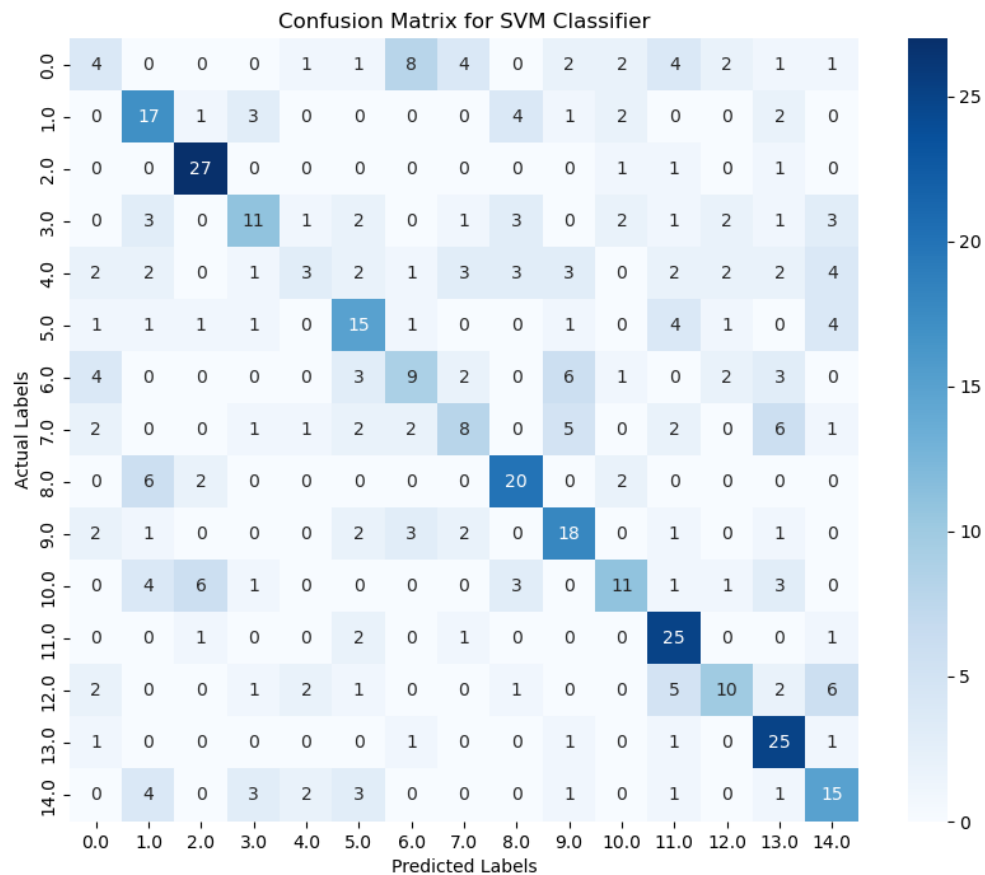
Overall it is great that the models can improve over random guessing when it comes to classification but an accuracy of only 50% is not very good. The models both have removed important spatial distribution information that could help with classification. In addition the models are ultimatelt linear in nature with some primitive human engineered features when it comes to descriptors. The SIFT descriptors are good for finding unique points and mapping them across images or frames but they are not very good at describing the overall image.