

Project 3

Boris Castillo

The purpose of this lab is to gain experience implementing a 'Bag of Visual Words' image classification method. As well as a little experimenting with deep learning.

The first step is to process our training images in order to obtain some sort of defining feature for each one. For this I used sift, finding key points for each image. Next we cluster the key points from all images.

Right Now each image is represented by its own set of key point descriptors. Our goal is to represent all images under a common set of descriptors, like a common language. The first step is to create that common language. In this case we use k-means to do so. The output is a common language with a pre-determined number of cluster centers. Next we must translate each image from its unique language to our common language. There are two ways of doing this. One is nearest neighbor and the other is local encoding.

In nearest neighbor key points are matched to one of the points from our common language. Now we create a histogram for each image of the key points in that image. Now we have a uniform signature for each image in a common language. Local encoding works similarly, but instead of assigning each key point to one cluster center, they are assigned to all cluster with some probability. The end result is the same, where each image now has a uniform signature which is described by our common language.

The next two steps are simple, create a model using our signatures and their known label.

The testing stage is similar, first we find the key points of each test image. Then we translated it to our common language. It's important that we do not create a new common language based on the test images, but use the one we have already created in the previous step. So we don't use k-means, and instead manually assign the nearest neighbor and local encoding clusters for each image.

Once we have the signatures for each test image, we can feed them through our models. Since every image is now described in the same common language the model should be able to correctly assign a label to each test image.

Part 2 and 3 are a different approach to the same task. The difference is that instead of creating our own common language we use an already established one. We still create a signature for each of our training and test images. The last step is the same, we create and use a model to predict labels.