Naive Pixel Classifier and Feature Extractor With CIFAR10 Dataset

Jawad Aziz Khan, ***ID - 1530457642***,
Ishfaq Zaman, ***ID - 1530173642***,
Kazi Habibur Rahman Dipto, ***ID - 1612823642***,
Sabit Ibn Ali Khan, ***ID - 1610377042***

1. Introduction

For Assignment 1, we have implemented 3 different albeit naive methods for extracting features and performing image classification using the CIFAR 10 image dataset. The techniques, or models, used are given below:
1) A naive classifier that flattens the image vector into a single dimension and matches based solely on the euclidean distance between each vector.
2) A naive classifier that extracts color histograms from images and matches using euclidean distance.
3) A model based around the Histogram of Oriented Gradients method.

Using the 4 techniques mentioned above, we used 1-NN classifier to compare features of images. When comparing the features, we found the nearest neighbor to each class and assumed the label of the nearest neighbor to be the label of the class as well, and this was our predicted result.
Obviously, each technique gave us varying results. Using color histograms gave us a result of 20 percent. On the other hand,when we converted raw pixel data to 1-dimensional vector, the result we got was 25 percent.

2. Literature Review

We have read several papers regarding image classification methods and feature extraction to better our understanding of complex computer vision algorithms. Below is a detailed review of some of the papers:

2.0.1. 1-NN

1-NN is a nearest neighbor classier and it is highly intuitive. In simple terms, to identify a certain class, we need to find its nearest neighbor among all the training points, and then assign to the class we are trying to identify the label of its nearest neighbor. 1-NN is conceptually very simplistic and it performs rather well in low dimensions for complicated decision surfaces. The metrics that have been used to measure the success of the 1-NN classifier are:
1) Confusion Matrix: Given a data set whose true values are known, a confusion matrix can be used to describe the performance of any classification methods, in this case 1-NN, on said data set.
2) Accuracy : Accuracy calculates how often our classifier predicts correctly.

Accuracy = True Positive + True Negative) / Total
3) Precision: Precision is simply a measure of how many correct ”yes” predictions our classification method gives us.
Precision = True Positive / (True Positive + False Positive)
4) Recall: Also known as sensitivity, is a measure of how many classes were identified correctly.
Recall= True Positive / (True Positive + False Negative)
5) ROC curve: It is a curve that visualizes the performance of any classifier method over all possible threshold values. The y-axis of the curve is True Positive rate and the x-axis is False Positive rate.
6) F1 score : The F1 score is used to measure the accuracy of any test. It takes into account the recall and precision values that we have talked about earlier. It is the weighted average of precision and recall. The best F1 score is 1, indicating perfect precision and recall, and the worst F1 sore is 0.

2.0.2. HOG

Histograms of Oriented Gradients for Human Detection is a paper written by Navneet Dalal and Bill Triggs. The purpose of the paper was to find a novel way of detecting the human form in a given image. It shows that HOG descriptors perform much better than other existing methods when it comes to human detection. The paper comes to a conclusion that factors such as good orientation binning and contrast normalization all play a big role in determining how good the results are.

The principle of the methodology is that the shape and appearance of an object in an image can be identified to a good enough extent by observing the gradient distribution of the intensity and by identifying the edges. In the HOG method, the image to be classified is taken and the gradient in a certain region is calculated and the gradients are categorized into various bins based on their orientations. To achieve invariance to factors such as differences in lighting, shadows and so on, contrast normalization is carried out on each block.

One advantage of the HOG method of classification is that it is able to achieve an acceptable amount of rotational and scale invariance as well as other forms of invariance. The paper used two data sets, one being the MIT pedestrian database (200 test images and 509 training images). The range of different poses of the pedestrians in this data set were very limited and as a result the detectors gave almost flawless results when using this data set. To make it more challenging for the detectors, a more complicated data set, ‘INRIA’ (1805 images) was used. The people in these photos were cropped from a diverse amount of personal photos and the poses and background the people were standing against largely varied.

2.0.3. SIFT

Object Recognition from Local Scale-Invariant Features is a paper authored by David G. Lowe. The detector system mentioned in the paper relies on image features that are unaffected by mane factors.The system uses local factors that are invariant to factors such as rotation and transformation. The features are identified by detecting certain stable points in the image. Keys are made from the images and said keys are sent to a nearest-neighbor detecting system that identifies keys that match to candidates. Previous methods of object detection such as color histograms, while having shown efficiency on isolated objects , are not very effective on images that are cluttered or occluded due to their features being much more global in nature

The SIFT method breaks down an image into a number of local features that are invariant to certain factors. This system is more efficient when compared to previous systems because image features generated by previous systems were variant to changes in scale and rotation. SIFT detects the local image features by initially identifying that are either the maximum or a minimum of a difference-of-Gaussian function. Each local image feature becomes partially variant to affine or three dimensional projection by blurring the locations of the gradients in the image. The vectors produced are called SIFT keys. A single image will produces on an order of 1000 keys, needing only 1 second of computation time to do so.

The generated keys are used as the input of a nearest-neighbor method to match the keys to candidate objects. A minimum of 3 keys must match with candidate objects for the conclusion to be drawn that the object might possibly be present there.

3. Experimental Setup

4. Results and Discussions

5. Conclusion