Unsupervised Feature Learning for Object Classification

Laxman Dhulipala, Harry Gifford, Wangzi He {ldhulipa, hgifford, wangzih}@andrew.cmu.edu

Contribution

We implement an unsupervised method for learning features from images, which are then used for image classification. Our method is robust, and we experimentally show its performance on the CIFAR-10 dataset, a standard dataset in the machine learning community.

Problem

Finding a robust and compact set of features is a fundamental problem in machine learning, with applications to a number of applications, including classification and similarity-detection. Finding a compact basis for representing images is therefore one of the first tasks faced by both academics and industry scientists working on image classification.

Motivation

Representing an $n \times n$ image using all n^2 pixels is almost always highly-wasteful of computational resources. This is due to two neighboring pixels being highly correlated, as they are likely to have very similar RGBA values. Learning more general features allows us to represent data in a much more efficient manner, and reduces both the memory and space utilization of machine learning algorithms.

Methodology

Our approach involves steps to learn a feature representation. We are going to explain each step in the following sections.

Preprocessing

1. We extract random patches (8 × 8) from unlabeled training images.
2. We apply PCA on those patches to reduce feature dimensions since each pixel is a feature in the original patch. (Optional step)
3. Normalize data so that mean is zero and variance is one.
4. We whiten data to eliminate correlations between neighboring pixels. (Neighboring pixels tend to have similar colors.)

Clustering

We run K-Means on the preprocessed patches, using a predetermined parameter k. We want to select k to be as large as possible without having many singular clusters where a singular cluster is a cluster with no elements or very few elements in it. Clusters are computed using the l_2 norm, although other metrics such as cosine-distance can easily be used. [1]

Centroids as Features

Learned features show noticable similarities to Gabor wavelets, and patterns exhibited by neurons in the V1 cortex of the brain.

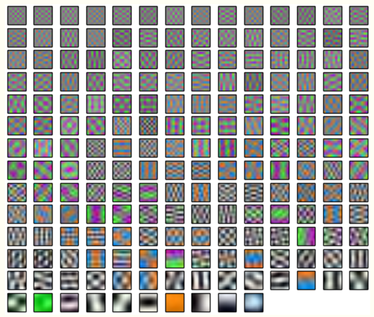

Figure 1: All 192 eigenvectors from a selection of 8×8 patches.

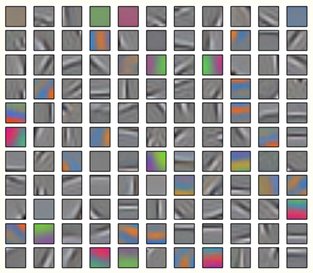

Figure 2: 100 centroids learnt from applying K-means to 8×8 patches.

Pooling

1. For each labeled image, we extract all possible patches for it.
2. Each patch extracted will correspond to a K-dimension vector. Let the vector corresponding to the ith patch be v_i. Element v_{ij} equals max(0, μ_i - x_{ij}) where μ_i is the average distance from the patch to each centroid and x_{ij} is the distance from patch i to centroid j.
3. We divide those patches into n groups and sum over each single column of the data in each group. We will obtain n × K values and those will be the new features. Finally, images are classified as follows:

Classification

1. Divide labeled data into training and test sets.
2. Pass these new features into SVM or other classifiers.

Results

The method was evaluated on the GavabDB expression dataset which contains 427 Scans, with 3 neutral scans and 4 expression scans per ID. To test the impact of expression invariance on neutral data we used the UND Dataset from the Face Recognition Great Vendor Test, which contains 953 neutral scans with one to eight scans per subject.

Expression neutralization improves results on the expression dataset without decreasing the accuracy on the neutral testset. Plotted is the ratio of correct answers to the number of possible correct answers.

Plotted are precision and recall for different retrieval depths. The lower precision of the UND database is due to the fact that some queries have no correct answers.

Open Questions

While the expression and identity space are linearly independent, there is some expression left in the identity model. This is because a "neutral" face is interpreted differently by the subjects. We investigate the possibilty to build an identity/expression separated model without using the data labelling, based on a measure of independence.

References

[1] A. Coates and A. Y. Ng. Learning feature representations with k-means. In *Neural Networks: Tricks of the Trade*, pages 561–580. Springer, 2012.