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

We want to learn feature representation for data with arbitrary form when the number of unlabeled data is much more than the number of labeled data.

Motivation

Learning features allows us to represent data in a more efficient manner.

Methodology

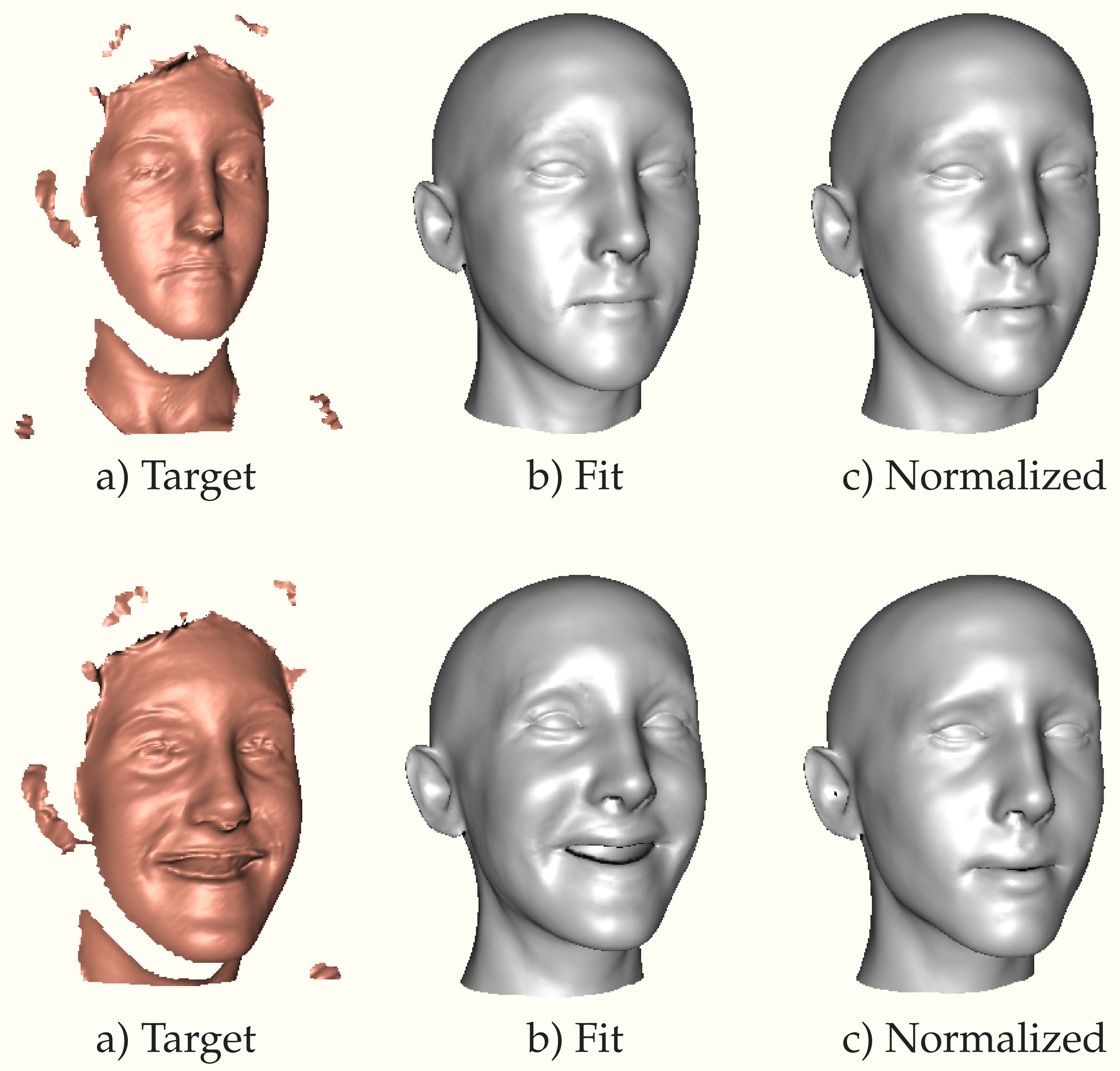
Our approach involves a few steps to learn a feature representation. We are going to explain each step in the following sections.

Preprocess

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. Parameter K is predetermined. We want to select K to be as large as possible without having lots of singular clusters. Singular cluster is a cluster with no elements or very few elements in it. We use euclidean distance metric. After clustering, we will obtain K centroids. [?]



Plots for K-Means and Sparse K-Means.

Clustering

Selection of K is a function of how much data we have and runtime of the algorithm.

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.

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

References

[1] A. Coates, A. Y. Ng, and H. Lee. An analysis of single-layer networks in unsupervised feature learning. In *International Conference on Artificial Intelligence and Statistics*, pages 215–223, 2011.

Funding

This work was supported in part by Microsoft Research through the European PhD Scholarship Programme.