Learning Attribute Spaces for One-Shot Recognition

Kevin Tang, Cornell University
Marshall Tappen, University of Central Florida
Rahul Sukthankar, Intel Research/CMU

Abstract

Humans can readily identify new object classes when shown just a single image of a new object. In the field of object recognition, we term this as one-shot recognition. Our goal is to enable a machine to perform one-shot recognition at a competitive level by utilizing novel machine learning ideas.

Background Information

•Current methods for object recognition consist of giving the system lots of training (labeled) images to learn from.

•We would like to minimize manual labor involved in labeling images by requiring just one labeled image for each class.

Our Solution

•We use the idea of transfer learning, which consists of learning priors from a set of different, but related classes for which we have lots of training data for.

•This idea is formulated in our method of learning projections to a lower dimension.

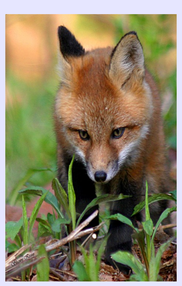
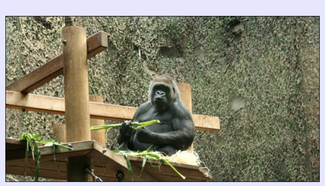
Image Dataset

•The image dataset that we use is the “Animals with Attributes” dataset, collected by Christoph Lampert (MPI).

•The dataset consists of images of 50 different animals that are presumed to share similar attributes.

•We split the animal classes into 40 training classes and 10 testing classes.

Example Images




Learning Projections

•We seek to learn a projection from the original feature space to a low dimensional space using the training classes.

•Our goal is to have animals with similar attributes grouped close to each other in this space.

•The projection is formulated as an optimization problem and minimized using stochastic gradient descent.

$$p_s = \sum_{i \in Te(s)} \frac{\exp(||\mathbf{o}(\mathbf{x}_i) - \mathbf{o}(\mathbf{x}_{Tr(s,C(i))})||^2)}{\sum_{j=1}^{N_C} \exp(||\mathbf{o}(\mathbf{x}_i) - \mathbf{o}(\mathbf{x}_{Tr(s,j)})||^2)},$$

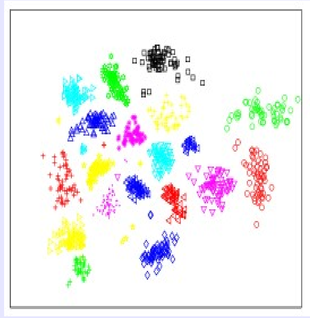


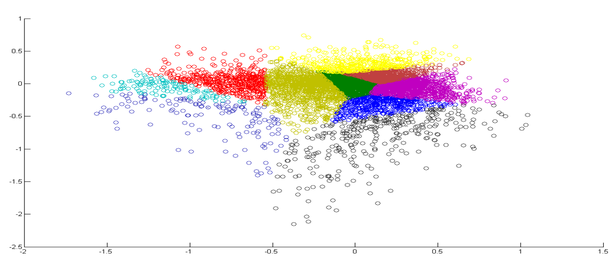
Image courtesy of Jacob Goldberger

Experimental Setup

•We devised two methods of evaluation to test our performance.

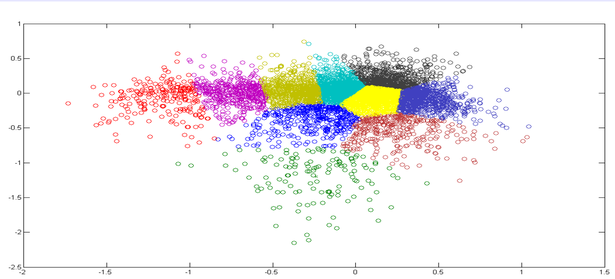
One-Shot Classification

•Using the single training image, we classify other randomly selected testing images using K-Nearest Neighbors.



Clustering

•Using equal numbers of images from each class, we cluster using K-means into 10 clusters, and assign labels based on cluster.



Results

•We compare our results to accuracies obtained by Lampert's method.

•Lampert's method incorporates additional semantic attribute knowledge (human measures of attributes for the animals), and so our problem formulation is more difficult than his.

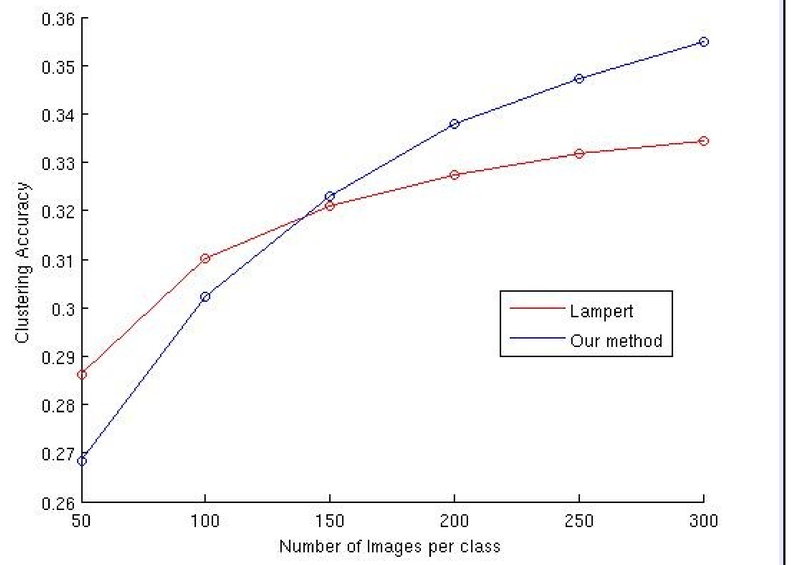
One-Shot Classification

•Chance: 10% accuracy

•Lampert: ~29.09% mean accuracy, ~51% max accuracy

•Our method: ~27.41% mean accuracy, ~51% max accuracy

Clustering



Conclusions

•We have introduced a novel formulation for performing transfer learning, and applied it to the task of one-shot recognition.

•With no additional semantic attribute knowledge, we are able to perform comparably with Lampert in the one-shot classification test, and outperform him in the one-shot clustering task.

Acknowledgements

•We would like to acknowledge Christoph Lampert for making available to us his code and data results.