Efficient Label Collection for Unlabeled Image Datasets

Maggie Wigness, Bruce A. Draper, J. Ross Beveridge
Department of Computer Science, Colorado State University

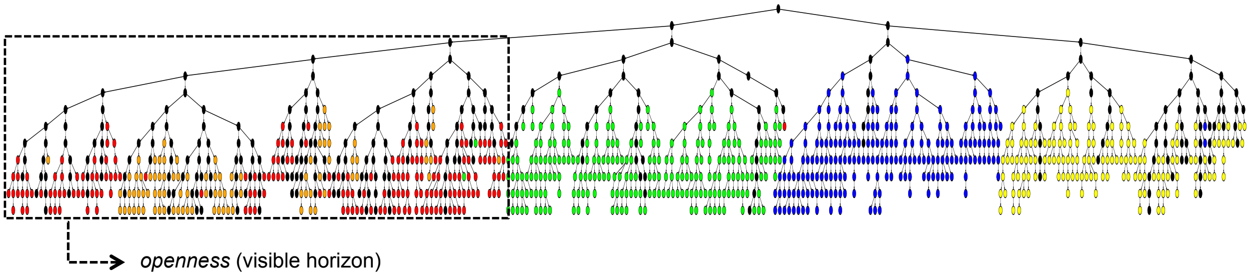

Figure 1: Hierarchical clustering to encode multiple visual concept granularities.

Visual classifiers are part of many applications including surveillance, autonomous navigation and scene understanding. To learn how to recognize concepts under common variations such as color, perspective or occlusion, visual classifiers often need large amounts of training data. The raw data needed to train supervised classifiers is abundant and easy to collect but contains no label information. Label collection requires significant human effort especially as the number of visual concepts in the data increases. Many real-world applications, however, do not have the resources needed to hand label all available training data.

Thus, efficient labeling schemes have emerged to reduce the labeling workload, while still producing sets of labeled data capable of training high performing classifiers. In the context of this paper, efficiency is defined relative to hand labeling each image in the training set individually. Two techniques are commonly used to reduce human workload. Active learning [4, 5] selects a subset of the most informative training data to label. Group-based labeling [1, 6] partitions the training data into groups and assigns a single label to a group, thereby labeling multiple images simultaneously.

Besides efficiency, interaction time and label accuracy are important to consider when evaluating labeling frameworks. Interaction time is higher for active learning frameworks since classifier re-training introduces latency between each labeled instance, but excessive latency may not be feasible in real-world labeling scenarios. Label accuracy is important because classifiers need reliable training data to learn from. Label accuracy decreases during group labeling when a group contains data from multiple visual concepts. In fact, it has become such a problem that group-based labeling techniques have recently begun to introduce additional labeling effort or latency to collect noise-free labels [3, 7, 8].

This paper is motivated by the hierarchical semantics that visual data represent, and demonstrates the benefits of exploiting this information to group and label data. Existing techniques may introduce label noise when feature patterns represent visual concepts broader than the classifier label set. For example, the 13-Scenes dataset [2] includes thirteen scene classes in addition to concepts like *outside* and *inside* scenes, or scenes that have an *openness* property, i.e., visible horizon, versus *closed/cluttered* scenes. Existing group-based labeling approaches assume groups will map to one of the thirteen scene classes when they may more accurately map to a coarser-grained concept.

Figure 2: *Openness* similarity shared by *coast* and *highway* classes.

We introduce a group-based labeling approach that uses hierarchical clustering to establish a space of groupings that span a spectrum of visual concept granularities. Figure 1 illustrates this grouping using five scene classes. Each color denotes which of the five scenes a group maps to, and black nodes indicate groups that represent multiple classes. Even for this simple example, grouping data by the five distinct scenes is not a trivial task. High intra-class similarity can be seen between red (*coast*) and orange (*highway*) nodes. This section of the hierarchy (dotted outline) more generally encodes the coarse-grained concept of *openness*. Since we do not know which groups in the hierarchy map to classifier-specific concepts, we maintain the hierarchy and use it to search for a set of groups that summarizes all possible candidate visual concepts in the data.

For efficiency, we only label groups that likely represent a different visual concept than their ancestors. In other words, we search for locations in the hierarchy where groups transition from coarser-grained to finer-grained concepts. We use large structural changes in local neighborhoods of the hierarchy to indicate a change in visual concept. Specifically, we suggest that the dominate direction of variance of a group of data provides information about the underlying structure and its corresponding concept. When there is a change in the direction of variance, the underlying concept of a group may also change. Labels are assigned with hierarchical semantics in mind, producing a set of labeled groups collected efficiently with minimal label noise. This ultimately results in the ability to train higher performing classifiers with less labeling effort than existing techniques, which is demonstrated using benchmark data.

Further, our framework hierarchically clusters training data only once, eliminating latency during the labeling process. This keeps the human interaction time to a minimum and makes our labeling system real-world feasible. We demonstrate the speed of our labeling system in a real-world scenario, where a single human annotator labeled new training data in less than 45 minutes for an autonomous navigation task.

[1] D. Dai, M. Prasad, C. Leistner, and L. Van Gool. Ensemble partitioning for unsupervised image categorization. In *Proc. ECCV*, 2012.

[2] L. Fei-Fei and P. Perona. A bayesian hierarchical model for learning natural scene categories. In *Proc. CVPR*, 2005.

[3] C. Galleguillos, B. McFee, and G. G. Lanckriet. Iterative category discovery via multiple kernel metric learning. *IJCV*, 2014.

[4] A. J. Joshi, F. Porikli, and N. Papanikolopoulos. Multi-class active learning for image classification. In *Proc. CVPR*, 2009.

[5] X. Li and Y. Guo. Adaptive active learning for image classification. In *Proc. CVPR*, 2013.

[6] D. Liu and T. Chen. Unsupervised image categorization and object localization using topic models and correspondences between images. In *Proc. ICCV*, 2007.

[7] M. Wigness, B. A. Draper, and J. R. Beveridge. Selectively guiding visual concept discovery. In *Proc. WACV*, 2014.

[8] C. Xiong, D. M. Johnson, and J. J. Corso. Spectral active clustering via purification of the *k*-nearest neighbor graph. In *Proc. ECDM*, 2012.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.