Supplementary material for paper “Multi-Level Active Prediction of Useful Image Annotations for Recognition”

1 Introduction

This supplementary document contains example images to illustrate the datasets used in our experiments and the complete set of learning curves for all classes in the two datasets.
Figure 1: Sample images from the SIVAL dataset. Each image is treated as a bag of segments and positive bags are images that contain the object, while negative bags are randomly sampled from all other categories. This figure shows example images for two of the 25 categories in the dataset.
Figure 2: Sample images from the Google dataset. Images are downloaded from the web using the category name as the keyword, and each returned set is treated as a positive bag. Negative bags are taken from all other categories. This figure shows an example positive and negative bag for two of the seven categories in this dataset.
Area under ROC

Category − ajaxorange

Category − apple

Category − banana

Category − blueshirt

Category − candelehanger

Category − cardboardbox

Category − checkeredscarf

Category − cokcan

Category − dataminingbook

Category − dirfryrunner

Category − dirfewdog

Category − fabricpatternbox

Category − feltsewing

Category − glassbrush/dust

Category − goldmedal

Category − greasebox

Category − julespot

Category − largespoon

Category − rapbook

Category − smileyfacedoll

Category − sprinbank

Category − stripeheadsbook

Category − translucentbook

Category − wad40can

Category − woodrunglass

Figure 3: Learning curves for our multi-level approach and three baselines on all 25 classes in the SIVAL dataset. All results are averaged over five random trials. Our Multi-level approach shows more gains per unit cost than the baselines on most of the categories. Due to space restrictions, in the main paper submission, we could show only three representative categories in Figure 2(a). This figure gives the remainder.

4
Figure 4: Learning curves for our multi-level approach and three baselines on the entire Google dataset consisting of seven categories. All results are averaged over five random trials. Our approach shows better gains for most of the categories. Multi-level random outperforms single-level random on this dataset, which we attribute to the larger number of positive bags chosen and the larger gain per addition. On the Airplane category, the curves for multi-level and single-level active are similar because it happened that the gains in VOI on adding positive bags did not overcome the higher cost of labeling them, and so instances were chosen all the time.