ConceptLearner: Discovering Visual Concepts from Weakly Labeled Image Collections

Bolei Zhou[†], Vignesh Jagadeesh[‡], Robinson Piramuthu[‡]
[1]MIT. [2]eBay Research Labs.

Discovering visual knowledge from weakly labeled data is crucial to scale up computer vision recognition systems, since it is expensive to obtain fully labeled data for a large number of concept categories. In this paper, we propose ConceptLearner, which is a scalable approach to discover visual concepts from weakly labeled image collections. Thousands of visual concept detectors are learned automatically, without human in the loop for additional annotation. We show that these learned detectors could be applied to recognize concepts at image-level and to detect concepts at image region-level accurately. Under domain-specific supervision, we further evaluate the learned concepts for scene recognition on SUN database and for object detection on Pascal VOC 2007. ConceptLearner shows promising performance compared to fully supervised and weakly supervised methods.

The contributions of this paper are as follows:

- scalable max-margin algorithm to discover and learn visual concepts from weakly labeled image collections.
- domain-selective supervision for application of weakly-learned concept classifiers on novel datasets.
- application of learned visual concepts to the tasks of concept recognition and detection, with quantitative evaluation on scene recognition and object detection under the domain-selected supervision.

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

Project page is at http://conceptlearner.csail.mit.edu/

Figure 1: **Discovered Concepts:** Illustration of learned concepts from NUS-WIDE and SBU datasets. Each montage contains the top 15 positive images for each concept, followed by a single row of 5 negative images. 4 sub-category concept detectors for car and boat respectively are illustrated in (a), based on concepts learned from NUS-WIDE. The title shows the name set for each concept from NUS-WIDE. Phrases for SBU dataset are shown in titles as in (b). We use each tag/phrase to represent a concept and group the associated images together. Within each such group (say, cat, boat), we group images based on visual features only, as we want to have visually similar cluster for one concept. Label vectors (say, car-racing-race, car-automobile-truck, car-automobile-vehicle, car-road-light) are further used to name these clusters after hard instance learning using ConceptLearner. This refined collection of groups is then used to learn concept classifiers and detectors. Examples of positive and negative samples for few such concept classifiers are shown in this figure.

Figure 2: **Concept Detection:** Results of concepts discovered from (a) NUS-WIDE and (b) SBU. Top 20 bounding boxes with high detector responses are shown. Note that for legibility we manually overlaid the text labels with large fonts.

Figure 3: **Concept Recognition:** Illustration of concept recognition using concepts discovered from (a) NUS-WIDE and (b) SBU datasets. Top 5 and 15 ranked concepts are shown respectively. These predicted concepts well describe the objects, the scene contexts, and the activities in these images.

Table 1: Comparison of methods with various kinds of supervision on Pascal VOC 2007. NUS-WIDE has missing entries since some object classes don't appear in the original tags. If we pool the concept detectors learned from SBU dataset and concept detectors learned from NUS-WIDE dataset together as initial concept pool, the final mAP is **23.2**.

Method	Supervision	mAP
SBU	Selected	**20.9**
NUS-WIDE	Selected	-
CVPR'14 (Divvala *et al*)	Webly	17.2
ECCV'12 (Prest *et al*)	Video	-
ICCV'11 (Siva *et al*)	Weakly	13.9
ICML'14 (Song *et al*)	Weakly	22.7
CVPR'14 (Girshick *et al*)	Full	44.7