
Weakly supervised learning from images and videos*

Cordelia Schmid

Abstract

With the amount of on-line available digital content growing daily, large-scale, weakly supervised learning is becoming more and more important. In this talk we present some recent results for weakly supervised learning from images and videos.

Standard approaches to object category localization require bounding box annotations of object instances. This time-consuming annotation process is sidestepped in weakly supervised learning, where the annotation is restricted to binary labels that indicate the absence/presence of object instances in the image. Our main contribution is a multi-fold multiple instance learning procedure, which prevents training from prematurely locking onto erroneous object locations. We also propose a window refinement method, which improves the localization accuracy by incorporating an objectness prior.

We, then, show how to move towards unsupervised discovery and localization of dominant objects from a noisy image collection of multiple object classes. The setting of this problem is fully unsupervised, without even image-level annotations or any assumption of a single dominant class. We tackle the discovery and localization problem using a part-based matching approach considering both appearance similarity and spatial consistency of candidate regions. Dominant objects are discovered and localized by comparing the scores of candidate regions and selecting those that stand out over other regions containing them.

Finally, we present work on learning object detectors from real world web videos known only to contain objects of a target class. We propose a fully automatic pipeline that localizes objects in a set of videos of the class and learns a detector for it. The approach extracts candidate spatio-temporal tubes based on motion segmentation and then selects one tube per video jointly over all videos.

*Machine Learning External Seminar, Gatsby Unit, 11 March 2015.