



Learning generative models of images and videos

Brendan J. Frey
University of Toronto, CA

Abstract

The best solutions to vision problems involve an unsupervised step where image pixels are converted into a descriptive code and a supervised step where the descriptive code is used to predict labels. The unsupervised step can be hand-coded (eg, as in region-specific SIFT histograms) or learned from training data (eg, as in PCA or deep belief networks). In this tutorial, I will discuss the state of the art in vision algorithms for several tasks involving the analysis of images and videos, but in the context of the question "Can unsupervised learning improve the performance of this method?" For example, I'll argue that commonly used approaches to obtaining the best performance on the PASCAL VOC challenge will fail to scale well to large numbers of object categories. I'll also review methods for learning generative models, including transformation- and color-invariant models, image epitomes, layered scene models, and hierarchical parts-based models.

Syllabus: Unsupervised Analysis of Images and Videos, Generative Models, Color-Invariant Models, Image Epitomes, Layered Scene Models, Hierarchical Parts-based Models