Predictive Vision

Carl Vondrick

Columbia University and Google Research, USA

Abstract

In this lecture, I will discuss our research to capitalize on large amounts of unlabeled video in order to train models for visual recognition. Leveraging millions of videos, our work develops methods for machines to learn perception tasks such as anticipating human actions in the immediate future, tracking visual objects, and recognizing ambient sounds. Our work takes advantage of the natural context available in video in order learn without human supervision, for example by transferring knowledge across modalities or generating videos of the future. Predictive vision provides a framework for efficiently learning from unlabeled data, enabling new applications in graphics, recognition, and robotics.