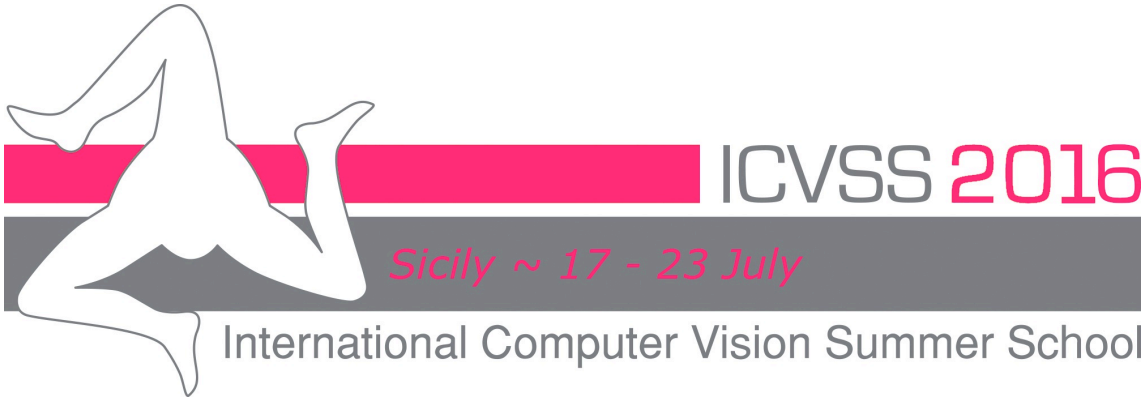

Images and Natural Language

Andrej Karpathy
Stanford University, USA

Abstract

Intelligent agents require the ability to perceive their environments, understand their high-level semantics, and communicate with humans. While computer vision has recently made great strides on visual recognition tasks, the predominant paradigm is to predict one or more fixed visual categories for each image. In this tutorial I will discuss recent advances that allow us to significantly expand the vocabulary of computer vision systems by treating natural language as a label space. In particular, I will describe the common tasks in this area such as image-sentence ranking, (dense) image captioning and visual Q&A, present recent state of the art approaches, and discuss current limitations and avenues for future work.