Visual Inference by Composition

Michal Irani

The Weizmann Institute of Science, Israel

Abstract

Inference tasks in Computer Vision often rely on having prior knowledge. This applies to Decision Problems (e.g., classification, detection, segmentation) as well as to Regression Problems (3D reconstruction, super-resolution, denoising, deblurring, etc.) Prior information is usually obtained by training on databases of examples (class-specific examples, or general 'natural image' databases), in order to extract informative parameters, features or models for classification, reconstruction, etc. While this approach has led to an impressive progress in Computer Vision in the past few years, it is restricted to a limited set of scenarios for which we can provide enough training examples with enough diversity in appearance.

In this talk I will show how complex visual inference tasks can often be performed without any prior examples, by ***exploiting data redundancy*** within and across different parts of the visual data. Comparing and integrating local pieces of visual information gives rise to ***complex notions of visual similarity*** and to a general ***"Inference by Composition"*** approach. This allows to make inferences about complex static and dynamic visual information, without training on prior examples.

I will explain the theory of "Inference by Composition" and how it can be approximated with computationally tractable methods. I will demonstrate the power of this approach to several different visual inference problems using images and video data. These will include (as time permits):

1. Detecting complex objects and actions in images and video (based on a single example, often only on a rough hand-sketch of what we are looking for).
2. Summarization of visual data (images and video).
3. Exceeding optical bounds of visual sensors by exploiting data redundancy (e.g., super-resolution).
4. Prediction of missing visual information (in images and video).
5. Detecting the "irregular" and "unexpected" (e.g., suspicious behaviors, defects, attention).
6. Segmentation of complex images ("Segmentation by Composition").
7. Separation of transparent layers.

If time permits, I will also talk about alignment and integration of information from different video sequences (sequence-to-sequence alignment).

Syllabus: Visual Inference, Space-time video analysis, Internal data redundancy, Object detection, Action detection, Visual summarization, Super-resolution, Visual prediction, Suspicious & salient behaviors, Segmentation, Transparency separation, Sequence-to-sequence alignment.