



ICVSS 2017

Sicily ~ 09-15 July

International Computer Vision Summer School

Training object localization models in weakly supervised settings

Vittorio Ferrari

University of Edinburgh, UK

Abstract

A key goal of computer vision is to interpret complex visual scenes, by recognizing visual concepts, localizing them, and understanding their interactions within the scene. To achieve this we need powerful visual learning techniques to acquire rich models capturing the diversity of the visual world. One crucial ingredient is the ability to learn visual localization models with minimal human supervision. This is necessary to scale to a large number of concepts and many training samples. In this lecture I will give an overview of weakly supervised techniques, that can learn without any location annotation, given only image-level labels. I will also talk about recent techniques in an intermediate regime where humans provide partial location information through answering simple questions, such as clicking in the middle of an object or verifying annotations produced by the learner. These seem to hit a particularly sweet spot in the trade-off between quality of the learned models and human supervision time.