

Abstract

“Learning high-level reasoning in and from images”

Prof Dr Christian Wolf, INSA, Lyon, France

Humans are able to infer what happened in a video given only a few sample frames. This faculty is called reasoning and is a key component of human intelligence. A detailed understanding requires reasoning over semantic structures, determining which objects were involved in interactions, of what nature, and what were the results of these. To compound problems, the semantic structure of a scene may change and evolve. In this talk we present research in high-level reasoning from images and videos, with the goals of understanding visual content (scene comprehension) or to make predictions of probable future outcomes, or to act in simulated environments based on visual observations. We present neural models addressing these goals through structured deep-learning, i.e. inductive biases in deep neural networks which explicitly model object relationships. We learn this models from data or from interactions between an agent and an environment.