

Visual Situation Recognition: An Integration of Deep Networks and Analogy-Making

This project investigates a novel approach to building computer systems that can recognize visual situations. While much effort in computer vision has focused on identifying isolated objects in images, what people actually do is recognize coherent *situations*—collections of objects and their interrelations that, taken together, correspond to a known concept, such as “a child’s birthday party”, or “a man walking a dog on the beach”, or “two people about to fight”, or “a blind person crossing the street”. Situation recognition by humans may appear on the surface to be effortless, but it relies on a complex dynamic interplay among human abilities to perceive objects, systems of relationships among objects, and analogies with stored knowledge and memories. No computer vision system yet comes close to capturing these human abilities. Enabling computers to flexibly recognize visual situations would create a flood of important applications in fields as diverse as medical diagnosis, interpretation of scientific imagery, enhanced human-computer interaction, and personal information organization.

The approach explored in this project integrates two previously studied approaches: brain-inspired neural networks for lower-level vision and cognitive-level models of concepts and analogy-making. In this integrated architecture, recognizing situations—via analogies with stored conceptual structures—will be a dynamic process in which bottom-up (perceptual) and top-down (conceptual) influences affect one another as perception unfolds. If successful, this system will be able to recognize visual situations in a way that scales well with the complexity of the scene and the abstract concept being recognized.

As part of this project, a number of benchmark image datasets—reflecting different abstract visual situations—will be collected to evaluate the recognition system. In addition, the PI will design and run a public competition on automated recognition of visual situations, using the collected datasets. This competition will spur research on this topic, and help researchers working in this area evaluate the success of various methods and gauge the current state of the art on abstract visual recognition. All source code and benchmarking databases developed in this project will be made publicly available via the web.