

A model of structure learning, inference, and generation for scene understanding

David Raposo

DeepMind, London, United Kingdom

Peter Dayan

Gatsby Computational Neuroscience Unit, London, United Kingdom

Demis Hassabis

DeepMind, London, United Kingdom

Peter Battaglia

DeepMind, London, United Kingdom

Abstract: Humans possess rich knowledge of the structure of the world, including co-occurrences among entities, and co-variation among their discrete and continuous features. But how people learn, infer and predict this structure is not well understood. Here we explore everyday scene understanding as a case study of people’s structural knowledge and reasoning. We introduce a probabilistic model over scene graphs that can learn the relational structure of objects and their arrangements and support inference and generation. Our model was able to learn the underlying structure of real-world scenes, and use it for inference and compression. In two human psychophysical experiments we found that a corresponding computational cognitive model was able to explain how people learn novel scene distributions and use it for classification and construction. Our work represents the first computational theory of human scene understanding that can account for people’s rich capacity for learning and reasoning about structure.