

Learning Spatial Common Sense with Geometry-Aware Deep Networks

Katerina Fragkiadaki, Carnegie Mellon University

Current state-of-the-art visual systems accurately detect object categories that are rare and unfamiliar for many of us, yet they neglect the basic principles of object permanence or spatial awareness that an one-year-old child has developed. In this talk, we will discuss models that marry two powerful ideas, geometry and deep visual representation learning, into neural architectures for mobile visual scene understanding that build and update a latent 3D feature map of the scene. We will show how these models learn to spatially reason, develop object permanence and accurate view predictions that generalize to truly novel scenes. We argue that latent 3D space-aware representations maybe a key to capturing Physics and spatial common sense for embodied agents that learn mental models by observing and moving in the world, as opposed to merely mapping pixels to labels using human supervision.