Visual Intelligence from Human Learning
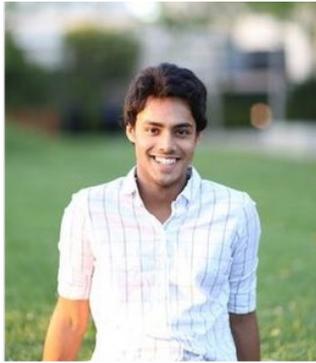
Wednesday, March 31, 2021
4:00 p.m.
Speaker: Ranjay Krishna

Zoom Presentation

Abstract: At the core of human development is the ability to adapt to new, previously unseen stimuli. We comprehend new situations as a composition of previously seen information and ask one another for clarification when we encounter new concepts. Yet, this ability to go beyond the confounds of their training data remains an open challenge for artificial intelligence agents. My research designs visual intelligence to reason over new compositions and acquire new concepts. My talk will explore these challenges and present the two following lines of work:

First, I will introduce scene graphs, a cognitively-grounded, compositional visual representation. I will discuss how to integrate scene graphs into a variety of computer vision tasks, enabling models to generalize to novel compositions from a few training examples. Since our introduction of scene graphs, the Computer Vision community has developed hundreds of scene graph models and utilized scene graphs to achieve state-of-the-art results across multiple core tasks, including object localization, captioning, image generation, question answering, 3D understanding, and spatio-temporal action recognition.

Second, I will introduce a framework for socially situated learning. This framework pushes agents beyond traditional computer vision training paradigms and enables learning from human interactions in online social environments. I will showcase a real-world deployment of our agent, which learned to acquire new visual concepts by asking people targeted questions on social media. By interacting with over 230K people over 8 months, our agent learned to recognize hundreds of new concepts. This work demonstrates the possibility for agents to adapt and self-improve in real-world social environments.

Bio: Ranjay Krishna is a 5th-year Ph.D. candidate at Stanford University, where he is co-advised by Fei-Fei Li and Michael Bernstein. His research lies at the intersection of computer vision and human-computer interaction; it draws on ideas from behavioral and social sciences to improve visual intelligence. His work has been recognized by the Christofer Stephenson Memorial award, as an Accell Innovation Scholar and by two Brown Institute for Media Innovation grants. His work has also been featured in Forbes magazine and in a PBS NOVA documentary. During his Ph.D., he re-designed Stanford's undergraduate Computer Vision course and currently also instructs the graduate Computer Vision course, Stanford's second largest course. He received an M.Sc. from Stanford University. Before that, he conferred a B.Sc. with a double major in Electrical Engineering and in Computer Science from Cornell University. In the past, he has interned at Google AI, Facebook AI Research, and Yahoo Research.

Yale University