

SenSIP Seminar Series

Leveraging Multimodal Embeddings for Zero-Shot Detection and Social Media Analytics

Presenter: Karan Sikka

Computer Vision Scientist, Center for Vision Technologies, SRI International

May 10 (Thursday), 2018, 3:00 PM / Room: GWC 487

Abstract

The recent progress in deep learning has resulted in performance improvements on several computer vision tasks. However, most deep learning algorithms still require massive amounts of (curated) labeled data for training. This is a very strong assumption and limits the applicability of these methods to specific problems. We propose to tackle these challenges by leveraging joint multimodal neural embeddings and show its application for two real-world problems. I will first discuss our recent work on a new task of zero-shot object detection- which targets the problem of detecting objects that were not seen during training. I will focus on the challenges associated with this task and two key ideas for solving them. In the second part of the talk I will focus on our work on social-media analytics and discuss our work on joint characterization of users and their posted content on social media platforms using multimodal embeddings. I will show that we are able to leverage millions of tweets and images to understand users interests across a large number of concepts- both concrete and abstract.

Biography:



Karan Sikka is a Computer Vision Scientist at Center for Vision Technologies, SRI International in Princeton, New Jersey. He graduated with a PhD degree from Machine Perception Lab at UCSD and was advised by Dr. Marian Bartlett. Before joining UCSD, he completed my bachelor's in ECE at Indian Institute of Technology Guwahati. His research interest in general spans joint multimodal analytics and computer vision problems related to classification and detection in both images and videos. During his PhD he primarily worked on problem related to action classification in videos for both recognizing human facial behavior and human actions. At SRI he have developed innovative prototypes and algorithms pertaining to deep multimodal (vision, language and audio) learning for understanding social media structure and content under the DARPA M3I, AFRL MESA and ONR CERROSS programs.

Refreshments

Sponsored by the SenSIP Center and NSF I/UCRC

Technical Co-Sponsorship by the IEEE Signal Processing and Communications Chapter, Phoenix Section

<http://engineering.asu.edu/sensip>

