



THE WEIZMANN INSTITUTE OF SCIENCE  
FACULTY OF MATHEMATICS AND COMPUTER SCIENCE

Vision and Robotics Seminar

Room 1 ,Ziskind Building  
on Sunday, Sep 08, 2019  
at 11:15

Special Seminar

Trevor Darrell  
UC Berkeley

Adapting and Explaining Deep Learning for Autonomous Systems

Abstract:

Learning of layered or "deep" representations has recently enabled low-cost sensors for autonomous vehicles and efficient automated analysis of visual semantics in online media. But these models have typically required prohibitive amounts of training data, and thus may only work well in the environment they have been trained in. I'll describe recent methods in adversarial adaptive learning that excel when learning across modalities and domains. Further, these models have been unsatisfying in their complexity--with millions of parameters--and their resulting opacity. I'll report approaches which achieve explainable deep learning models, including both introspective approaches that visualize compositional structure in a deep network, and third-person approaches that can provide a natural language justification for the classification decision of a deep model.