



THE WEIZMANN INSTITUTE OF SCIENCE
FACULTY OF MATHEMATICS AND COMPUTER SCIENCE

Vision and Robotics Seminar

Room 1 ,Ziskind Building
on Thursday, Nov 10, 2016
at 12:15

Yedid Hoshen
Hebrew University of Jerusalem

End-to-End Learning: Applications in Speech, Vision and Cognition

Abstract:

One of the most exciting possibilities opened by deep neural networks is end-to-end learning: the ability to learn tasks without the need for feature engineering or breaking down into sub-tasks. This talk will present three cases illustrating how end-to-end learning can operate in machine perception across the senses (Hearing, Vision) and even for the entire perception-cognition-action cycle.

The talk begins with speech recognition, showing how acoustic models can be learned end-to-end. This approach skips the feature extraction pipeline, carefully designed for speech recognition over decades.

Proceeding to vision, a novel application is described: identification of photographers of wearable video cameras. Such video was previously considered anonymous as it does not show the photographer.

The talk concludes by presenting a new task, encompassing the full perception-cognition-action cycle: visual learning of arithmetic operations using only pictures of numbers. This is done without using or learning the notions of numbers, digits, and operators.

The talk is based on the following papers:

[Speech Acoustic Modeling From Raw Multichannel Waveforms](#), Y. Hoshen, R.J. Weiss, and K.W. Wilson, ICASSP'15

[An Egocentric Look at Video Photographer Identity](#), Y. Hoshen, S. Peleg, CVPR'16

[Visual Learning of Arithmetic Operations](#), Y. Hoshen, S. Peleg, AAAI'16