מכון ויצמן למדע

THE WEIZMANN INSTITUTE OF SCIENCE

FACULTY OF MATHEMATICS AND COMPUTER SCIENCE

Vision and Robotics Seminar

Room 1 ,Ziskind Building
on Wednesday, May 25, 2016
at 11:15

JOINT SEMINAR WITH MACHINE LEARNING & STATISTICS

Bill Freeman
MIT

Visually Indicated Sounds

Abstract:

Children may learn about the world by pushing, banging, and manipulating things, watching and listening as materials make their distinctive sounds-- dirt makes a thud; ceramic makes a clink. These sounds reveal physical properties of the objects, as well as the force and motion of the physical interaction.

We've explored a toy version of that learning-through-interaction by recording audio and video while we hit many things with a drumstick. We developed an algorithm the predict sounds from silent videos of the drumstick interactions. The algorithm uses a recurrent neural network to predict sound features from videos and then produces a waveform from these features with an example-based synthesis procedure. We demonstrate that the sounds generated by our model are realistic enough to fool participants in a "real or fake" psychophysical experiment, and that the task of predicting sounds allows our system to learn about material properties in the scene.

Joint work with:
Andrew Owens, Phillip Isola, Josh McDermott, Antonio Torralba, Edward H. Adelson
http://arxiv.org/abs/1512.08512