LEARNING TO SEE AND HEAR WITHOUT HUMAN SUPERVISION

ABSTRACT
Imagine the sound of waves. This sound may evoke many memories of days at the beach. A single sound serves as a bridge to connect multiple instances of a visual scene. It can group scenes that ‘go together’ and set apart the ones that do not. Co-occurring sensory signals can thus be used as a target to learn powerful representations for visual inputs without relying on costly human annotations.

In my talk, I will introduce effective self-supervised learning methods that curb the need for human supervision. I will discuss several tasks that benefit from audio-visual learning, including representation learning for action and audio recognition, visually-driven sound source localization, and spatial sound generation. I will introduce an effective contrastive learning framework that learns audio-visual models by answering multiple-choice audio-visual association questions. I will also discuss critical challenges we face when learning from audio supervision related to noisy audio-visual associations, and opportunities to overcome these challenges using robust learning algorithms.

BIO
Pedro Morgado is a Ph.D. candidate in the Electrical and Computer Engineering department at the University of California, San Diego advised by Prof. Nuno Vasconcelos. Prior to UC San Diego, he earned both B.Sc. and M.Sc. degrees from Universidade de Lisboa, Portugal. He has also interned at Adobe Research and Facebook AI Research. His main research interests lie in computer vision and deep learning, focusing on multi-modal self-supervised learning. Pedro is the recipient of a 4-year graduate fellowship awarded by the Portuguese Science and Technology Foundation.

DETAILS
Thursday, March 4, 2021 at 4:00 pm CST.