UNIVERSITY OF CENTRAL FLORIDA

CENTER FOR RESEARCH IN COMPUTER VISION

Dr. Ajmal Mian

The University of Western Australia

"Adversarial attacks on deep learning, defence mechanisms and their use for network explainability "

Thursday, February 6, 2020 · 2:00PM · HEC 101

ABSTRACT

Deep learning is at the heart of the current rise of machine learning and artificial intelligence. However, deep models are vulnerable to adversarial attacks in the form of subtle perturbations to inputs leading to incorrect decisions, often with high confidence. In this talk, I will give a brief introduction to the methods for generating adversarial perturbations. I will discuss early defence mechanisms, including our work, for defence against such attacks. I will then discuss our method for generating the first ever attack on skeleton based human action recognition that also translates to the physical world. Following this, I will explain our Label Universal Targeted Attack (LUTA) that makes a deep model predict a specific target label for any sample of only a given source class with high probability. This is achieved by stochastically maximizing the log-probability of the target label for only the source class while suppressing leakage to the non-source classes. LUTA perturbations achieve high fooling rates on the large-scale ImageNet models, and transfer well to the physical world. Finally, I will demonstrate the use of LUTA as a tool for deep model autopsy. LUTA results in interesting perturbation patterns revealing the inner working of the deep models and the training process itself exposes the feature embedding space.

BIOGRAPHY

Ajmal Mian is a Professor of CS at The University of Western Australia. He has received two prestigious fellowships and several research grants from the Australian Research Council and the National Health & Medical Research Council of Australia. He was the West Australian Early Career Scientist of the Year 2012 and has received several awards including the Excellence in Research Supervision Award, EH Thompson Award, ASPIRE Professional Development Award, Vice-chancellors Mid-career Research Award, Outstanding Young Investigator Award, and the Australasian Distinguished Doctoral Dissertation Award. He is an Associate Editor of IEEE Trans on Neural Networks & Learning Systems, IEEE Trans on Image Processing and the Pattern Recognition journal. He was a General Chair of the Int Conf on Digital Image Computing Techniques & Applications (DICTA) 2019, General Chair of the Asian Conference on Computer Vision 2018, Program Chair of DICTA 2012 and Area Chair of WACV 2019, WACV 2018, ICPR 2016, ACCV 2014. Ajmal Mian has supervised 15 PhD students to completion and has published over 180 scientific papers. His research interests are in computer vision, machine learning, 3D point cloud analysis, facial recognition, human action recognition and video analysis.