UNIVERSITY OF CENTRAL FLORIDA

CENTER FOR RESEARCH IN COMPUTER VISION

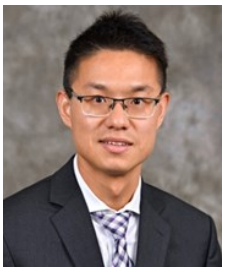
Dr. Chen Chen

University of North Carolina at Charlotte

"Adaptive and Efficient Deep Neural Networks for Edge Computing"

Thursday, March 25, 2021 · 11:00AM · CRCV Faculty Candidate

Meeting URL: https://ucf.zoom.us/j/95512185434?pwd=R1Bqd1FaV2Zzb2Zjc2t0cUlTNzc2Zz09&from=addon



ABSTRACT

Existing deep networks are mostly static - they can only run at a specific resource budget. In real applications, the resource constraints (e.g., computation and memory) are always changing with many conditions such as battery and workload. To meet the dynamic resource constraints, one needs to deploy several trained networks with different scales and switch among them, which consumes much more memory and is not efficient and scalable. This talk will discuss our recent work of addressing the important problem of efficient and adaptive inference under dynamic resource constraints. We propose a new training framework to mutually learn from network width and input resolution for adaptive accuracy-efficiency trade-offs. At runtime, our method dynamically selects the optimal resource-accuracy trade-off to fit the model's resource demand to the system's available runtime resources. This talk will conclude with a discussion of some future prospects for developing adaptive neural networks.

BIOGRAPHY

Dr. Chen Chen received the Ph.D. degree from the Department of Electrical Engineering, University of Texas at Dallas, Richardson, TX, USA, in 2016. He held a Postdoctoral Research Associate position at the Center for Research in Computer Vision (CRCV), University of Central Florida from July 2016 to June 2018. He joined the Department of Electrical and Computer Engineering at University of North Carolina - Charlotte as an Assistant Professor in 2018. His research interests include signal and image processing, computer vision, and deep learning. He published over 80 papers in refereed journals and conferences in these areas. He is an Associate Editor of Journal of Real-Time Image Processing and the IEEE JOURNAL ON MINIATURIZATION FOR AIR AND SPACE SYSTEMS.