**黃朝宗 (Chao-Tsung Huang)**
**副教授**
**國立清華大學電機工程學系**
chaotsung@ee.nthu.edu.tw

**Title**:
Bandwidth- and Energy-Efficient CNN Acceleration for Next-Generation Cameras and Displays

**Abstract**:
In the era of artificial intelligence, convolutional neural networks (CNNs) are emerging as a powerful technique for image processing, such as denoising, super-resolution, and even style transfer. They have shown great potential to bring next-generation cameras and displays to our daily life. However, it is difficult for conventional CNN accelerators to generate ultra-high-resolution videos at the edge due to their considerable DRAM bandwidth and power consumption. For example, the advanced FFDNet for denoising could demand 131 GB/s of DRAM bandwidth and 106 TOPS of computation at 4K UHD 30fps.
In this talk, I will introduce two of our recent works on tackling the challenges of DRAM bandwidth and power consumption. To begin with, I will talk about the possible image-pipeline revolution brought by image processing CNNs and elaborate their design challenges; for clarity, their differences from recognition CNNs will be highlighted. Then, I will present our first work—eCNN [MICRO'19]—which jointly optimizes inference flow, network model, instruction set, and processor design in a holistic way. In particular, a proposed block-based inference flow with hardware-oriented ERNet models can support FFDNet-level denoising and SRResNet-level super-resolution at 4K UHD 30 fps using less than 2 GB/s of DRAM bandwidth. Finally, I will introduce our second work—RingCNN [ISCA'21]—which exploits regular sparsity of ring algebra to achieve near-maximum hardware saving and graceful quality degradation for convolution engines. Layout results show that equivalent 41 TOPS of 8-bit computation can be delivered using 3.76 W and 2.22 W with 40 nm technology for sparsity at 50% (no quality drop) and 75% (0.11 dB of PSNR drop) respectively.

**Biography**:
Chao-Tsung Huang received the BS degree in electrical engineering and the PhD degree in electronics engineering from National Taiwan University in 2001 and 2005 respectively. He is now with National Tsing Hua University as an Associate Professor. Before joining NTHU, he was first with Novatek Microelectronics Corp. and then performed postdoctoral research at Massachusetts Institute of Technology and National Taiwan University. His research interests include computational photography, solid-state circuits, and computing architecture. He has published several research papers in prestigious conferences or journals of related fields: CVPR/ICCV/TPAMI (computer vision), ISSCC/VLSIC/JSSC (solid-state circuits), and ISCA/MICRO (computer architecture). He also serves as Associate Editor for IEEE TCSVT and Springer CSSP.
Dr. Huang was a recipient of the Young Scholar Innovation Award from Foundation For The Advancement Of Outstanding Scholarship (傑出人才基金會「年輕學者創新獎」) in 2020, the Outstanding Young Electrical Engineer Award from CIEE in 2019, the Outstanding Young Scholar Award from Taiwan IC Design Society (臺灣積體電路設計學會「傑出年輕學者獎」) in 2019, and the Junior Faculty Research Award from College of EECS, NTHU, in 2017.