Computer Vision – Efficiency, Hardware Friendliness and Applications

K.-T. Tim Cheng, HKUST

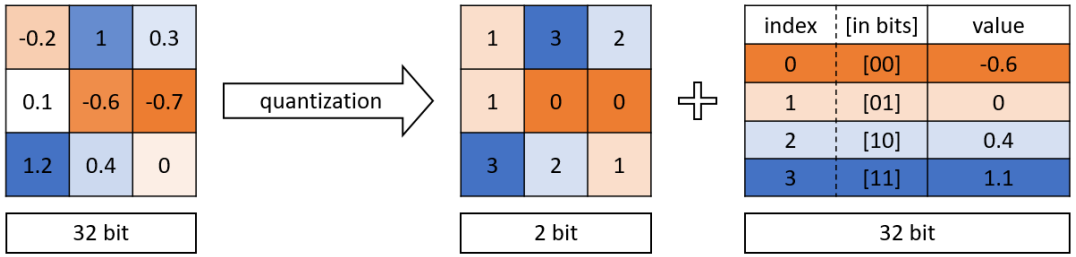
Vision & System Design Lab

- **Computer Vision**
  - Efficiency and Hardware Friendliness:
    - Neural network compression via quantization, channel pruning and neural architecture search (NAS)
  - Medical and AIoT Applications:
    - Medical image analysis – detection, segmentation, and diagnosis
    - Smart drone cinematography for filming human motion
    - Wellness analytics for seniors in elderly care centers
    - Food waste analytics in Canteens and visual feedback for behavior changes
- **AI Chips/Accelerators**

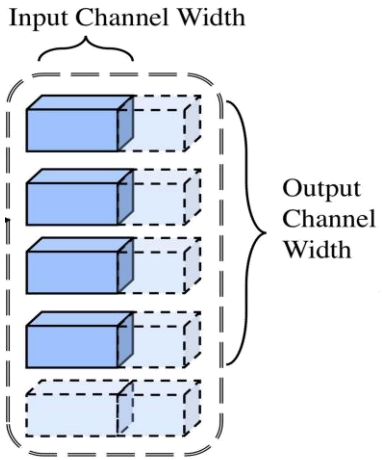
Neural Network Compression and HW implementation (AI Accelerator)

4 ways of compressing and accelerating the neural networks:

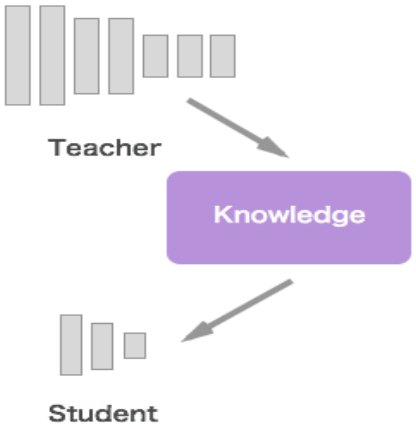
(1) Weights and activation quantization



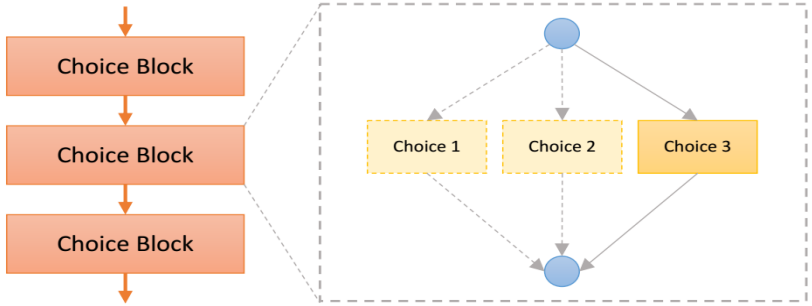
(2) Network channel pruning



(3) Knowledge distillation

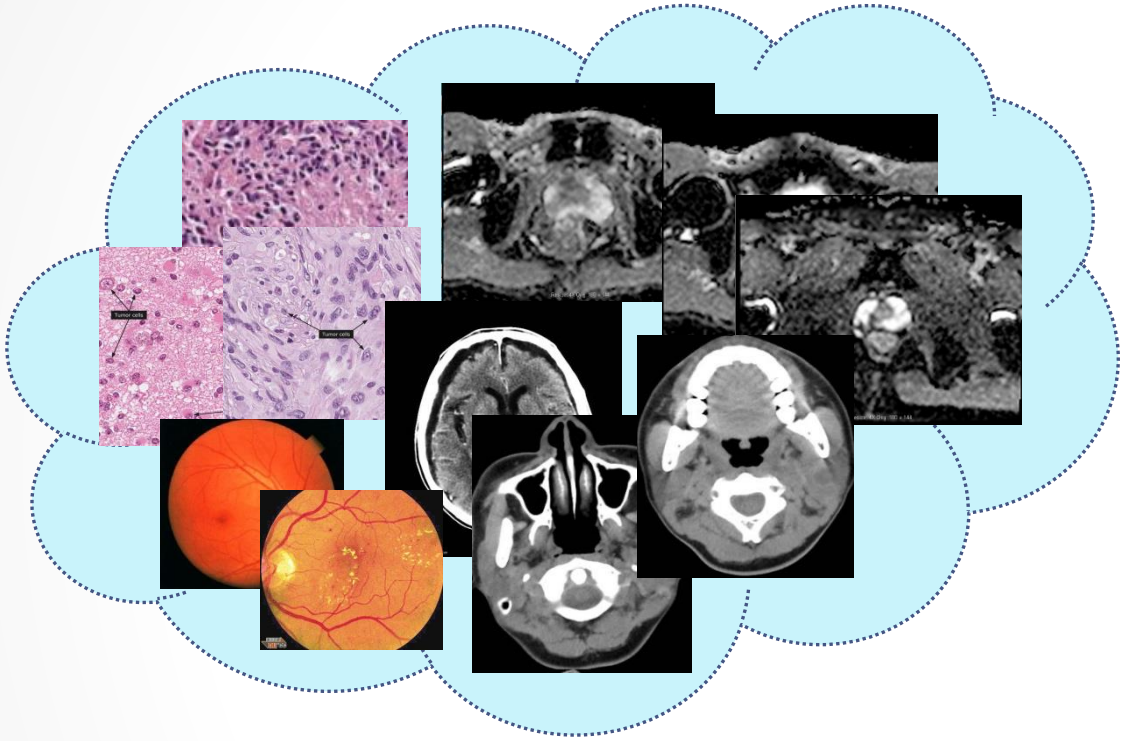


(4) Compact neural network design and search



- Quantization:
  - Bi-Real Net (ECCV 2018 & IJCV 2019)
  - ReActNet (ECCV 2020)
  - Binarizing MobileNet via Evolutionary Search (CVPR 2020)
- Network Channel Pruning:
  - MetaPruning (ICCV 2019)
- Neural Architecture Search:
  - Data-free Neural Architecture Search (submitted for publication)
- Knowledge distillation
  - Label Smoothing (ICLR 2021)

AI for Medical Imaging Aided Diagnosis



- Tremendous amount of Medical imaging data
  - 40 billion RMB in 2015, 80% of all medical data
  - 70% clinical diagnosis and treatment rely on imaging data
- Lack of expertise's for reading data (10-year growth cycle)

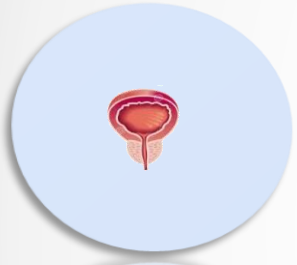
- **AI + Medical Imaging Analysis**

→

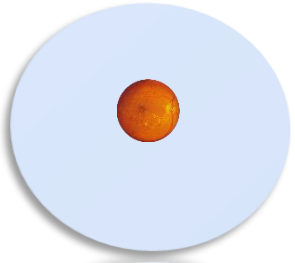
- ✓ **Objective and quantitative results**
- ✓ **Better efficiency and accuracy**
- ✓ **Optimized medical resources**

Our Research in AI for MID

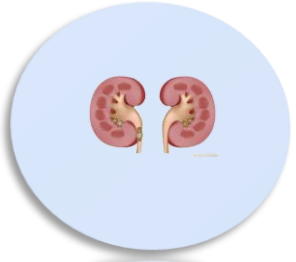
- **AI for Medical Imaging Diagnosis — 2nd Pair of Eyes of a Clinician**



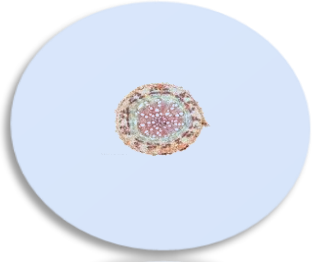
Prostate cancer detection & diagnosis in MRI



Cerebral aneurysm analysis in CT

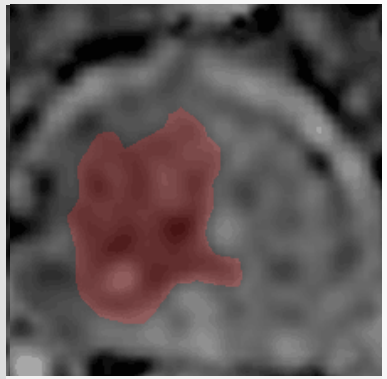


Renal function Analysis in DCE-MRI

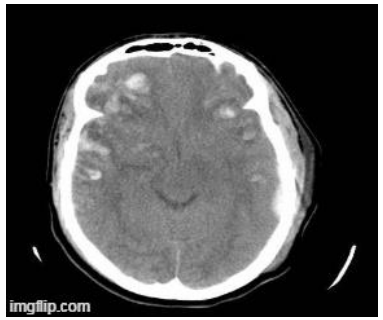


Large-Scale Histopathology Image Analysis

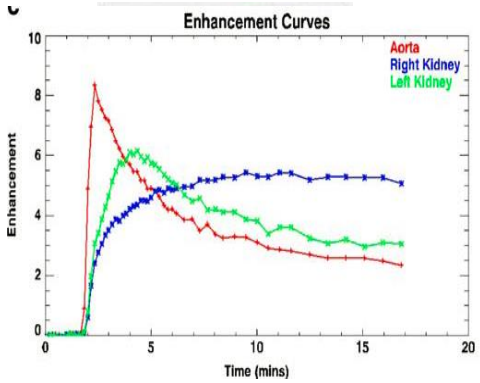
- Retinal Vessel Segmentation
- Gland Instance Segmentation
- Pancreas Segmentation
- Breast lesion detection
- ….

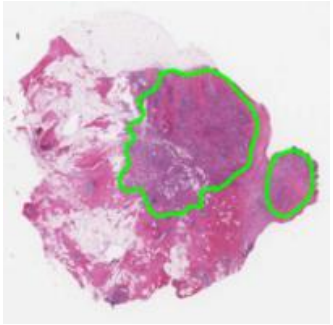


PCa segmentation



Aneurysm measurement & development analysis





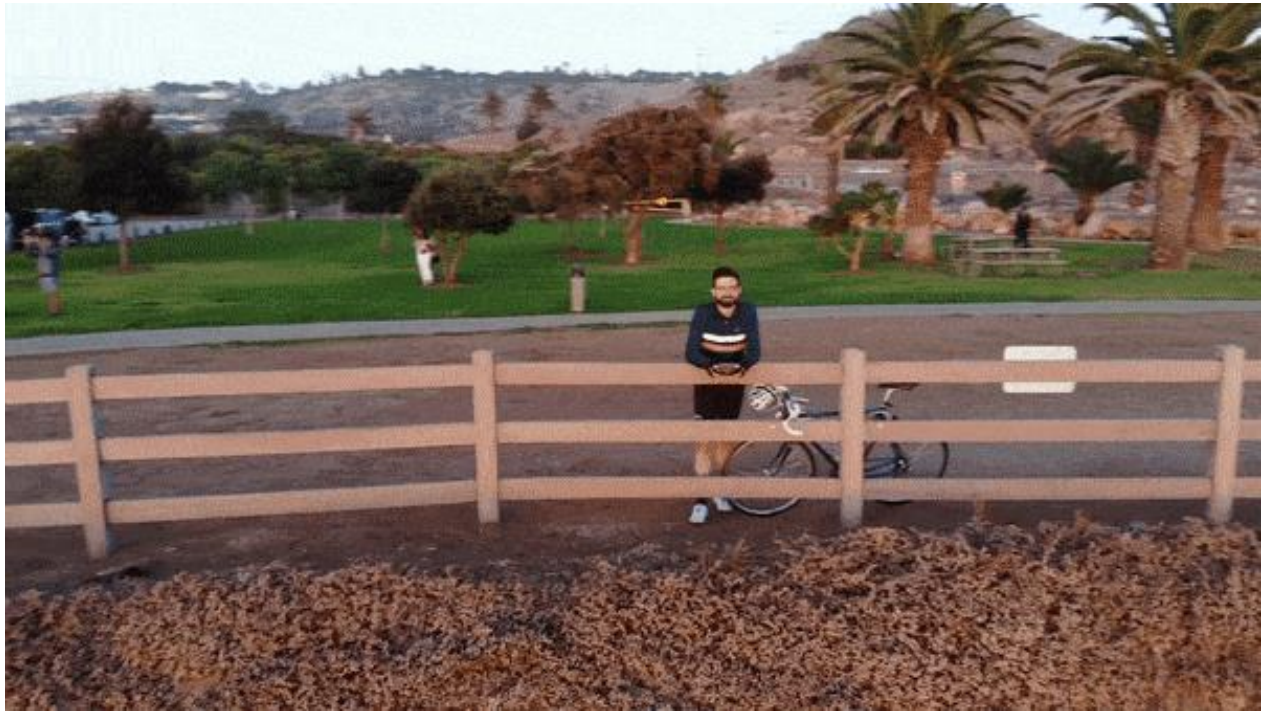
New techniques explored:

- Federated learning
- Domain adaptation
- Data augmentation
- ..

Vision & System Design Lab

- **Computer Vision**
  - Efficiency and Hardware Friendliness:
    - Neural network compression via quantization, channel pruning and neural architecture search (NAS)
  - Medical and AIoT Applications:
    - Medical image analysis – detection, segmentation, and diagnosis
    - **Smart drone cinematography for filming human motion**
    - Wellness analytics for seniors in elderly care centers
    - Food waste analytics in Canteens and visual feedback for behavior changes
- **AI Chips/Accelerators**

Existing Intelligent Applications in Consumer Drones



Captured by **DJI Mavic QuickShot mode**

1. Predefined camera trajectory
2. No camera-subject interaction

Captured by **DJI Mavic ActiveTrack mode**

1. Oversimplify the video content
2. Fixed viewpoint

Imitation Learning to Automate Drone Filming of Human Motion Video

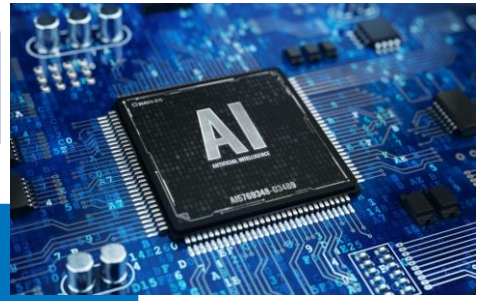
Our approach: Imitating given professional videos

Autonomously Captured Video with Style D

Vision & System Design Lab

- **Computer Vision**
  - Efficiency and Hardware Friendliness:
    - Neural network compression via quantization, channel pruning and neural architecture search (NAS)
  - Medical and AIoT Applications:
    - Medical image analysis – detection, segmentation, and diagnosis
    - Smart drone cinematography for filming human motion
    - Wellness analytics for seniors in elderly care centers
    - Food waste analytics in Canteens and visual feedback for behavior changes
- **AI Chips/Accelerators**

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY



AIR@InnoHK

ACCESS: AI Chip Center for Emerging Smart Systems

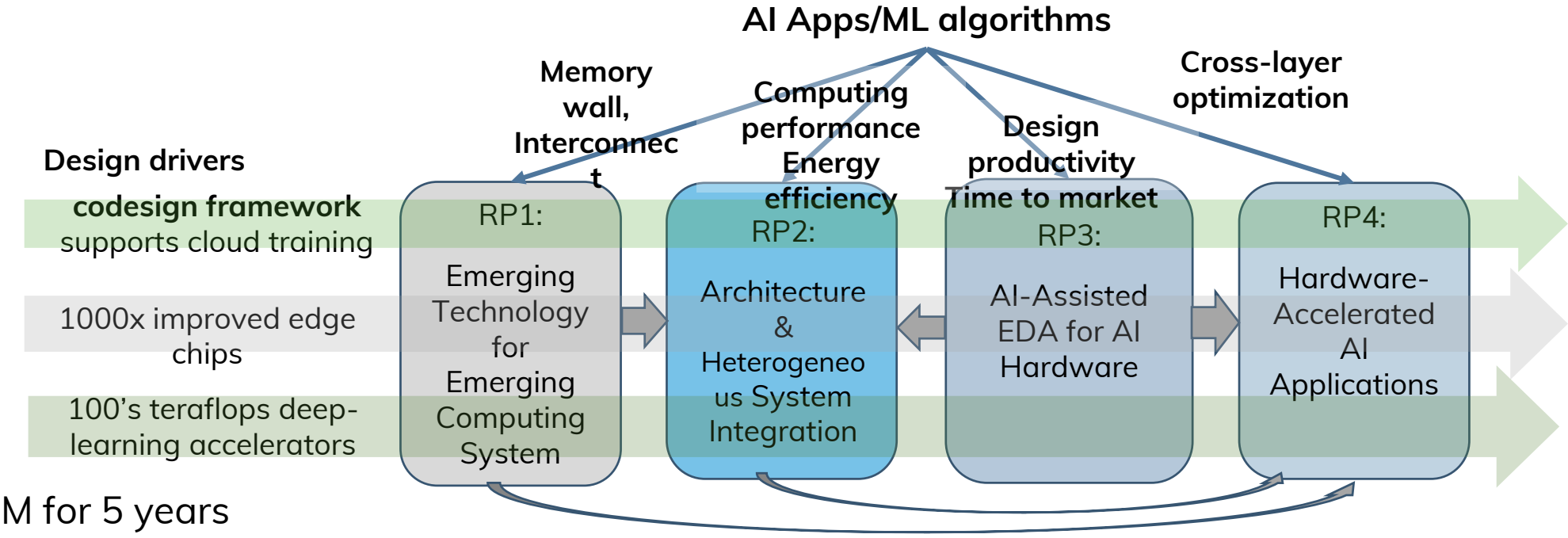
Lead University: HKUST

Participating Univ.: Stanford, UIUC, CUHK, HKU

Project Investigator Tim Cheng, HKUST

Mission: Realizing *ubiquitous AI applications in society by*:

1. Improving AI chip performance/energy efficiency by 1,000X
2. Accelerating AI chip design productivity and time-to-market



* Initial funding: $443.9M for 5 years

香港科技大學 THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

工學院 SCHOOL OF ENGINEERING

A Unique Team of World Leading Experts with Complementary Expertise

AIR@InnoHK

ACCESS: AI Chip Center for Emerging Smart Systems

IC Design

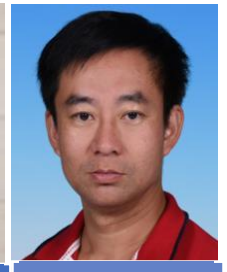
Tim CHENG
EDA, IC design and Computer Vision
- IEEE Fellow
- Former Director of MURI center on 3D hybrid circuits
- 11 best paper awards
- 12 US patents and co-founded 3 companies

Boris MURMANN
IC design
- IEEE Fellow
- Co-Director, Stanford SystemX Alliance
- Program Chair, ISSCC
- Director, Stanford's System Prototyping Facility

Deming CHEN
EDA, AI accelerator
- IEEE Fellow
- EiC, ACM Trans. on Reconfiguable Tech. & Systems
- 8 best paper awards
- ACM SIGDA Outstanding Faculty Award

Nanoelectronics & Emerging Devices

Philip WONG
VLSI, Nanoelectronics, emerging devices
- IEEE Fellow
- Co-Director, Stanford SystemX Alliance
- Executive Committee member of several multi-univ. centers



Mansun CHAN
Nanoelectronics
- IEEE Fellow
- Co-found >5 companies
- IEEE Electron Device Society Education Award 2017
- R&D 100 Award for the BSIM3 Project

Francois ANDRIEU
Nanoelectronics
- Led industrial projects of STMicroelectonics, Globalfoundries and IBM
- IEEE/SEE Brillouin Award, 2018

Design Methodology & EDA

Martin WONG
EDA
- IEEE and ACM Fellows
- Former Executive Associate Dean of Engineering, UIUC
- 6 best paper awards
- Graduated >50 PhDs

Evangeline YOUNG
Physical synthesis, EDA
- Founding Chair, IEEE CEDA Hong Kong Chapter
- Executive committee, Int'l conference on CAD 2019
- 4 best paper awards

Elyse ROSEBAUM
Nanoelectronics, circuits
- IEEE fellow
- Director, NSF Center for Advanced Electronics through Machine Learning
- SRC Technical Excellence Award

Architecture

CY TSUI
IC design, architecture & embedded systems
- Head, Division of Integrative Systems and Design
- PI for multiple ITF projects on sensor design and IC design

Wen-Mei HWU
Parallel computing
- IEEE and ACM Fellows
- ACM SigArch Maurice Wilkes Award
- ACM Grace Murray Hopper Award
- IEEE CS Charles Babbage Award

Priyanka RAINA
AI Accelerator, Architecture
- Research Scientist, Nvidia Research 2018
- Program Committee member and Session Chair, Hot Chips 2019

Embedded & Reconfigurable

Jiang XU
Architecture, silicon photonics
- Associate Editor, IEEE Trans. on CAD, Trans. on VLSI
- IEEE Distinguished Lecturer
- ACM Distinguished Speaker

Wei Zhang
EDA, reconfigurable system
- Associate Editor, IEEE Trans. on CAD, IEEE Trans. on VLSI, and Trans. on Reconfigurable Tech and Systems
- Best paper award, ICCAD 2017

Ngai WONG
EDA, machine learning
- Associate Editor, IEEE Trans. on CAD
- Co-founder, IEEE CEDA HK Chapter

AI/Machine Learning & Applications

David TSE
Information theory
- US NAE member
- IEEE Fellow
- IEEE Claude E Shannon Award 2017
- IEEE Richard W Hamming Medal 2019

Stephen BOYD
Optimization for ML, circuit design, etc.
- US NAE member
- Chinese Academy of Engineering foreign member
- IEEE Fellow
- Authors of several textbooks and open source SW packages

Edmund LAM
Imaging systems
- IEEE Fellow
- OSA Fellow
- IBM Faculty Award
- Published >300 papers
- Associate Dean of Engineering, HKU