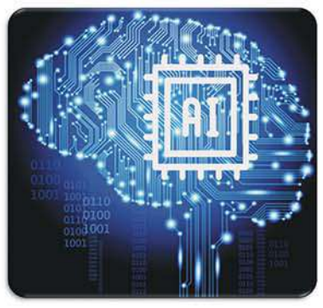
2022「中技社科技獎學金」

2022 CTCI Foundation Science and Technology Scholarship

研究獎學金

Research Scholarship



High-Efficient Learning and Inference Mechanisms for Brain-inspired Hyperdimensional Computing

國立台灣大學 電子工程學研究所 博士班四年級 莊育權

指導教授：吳安宇 教授

Motivation & Background

With the emergence of the Internet of Things (IoT), massive data streams are generated by IoT devices. Machine learning (ML)-on-edge provides a low-bandwidth and speedy way to process vast amounts of data from various tasks. However, processing high-complex DNNs on IoT devices poses substantial technical challenges due to the limited hardware resources. Recently, brain-inspired hyperdimensional computing (HDC) has gained much attention from academia and industry due to its (i) highly efficient computing, (ii) robustness against hardware failures, and (iii) fast/few-shot learning capability. These technical advantages make HDC suitable for efficient computing on edge devices. In this research, we aim to exploit the great potential of HDC and further increase its learning capability and computing efficiency for today's IoT systems.

Challenges of On-Device Learning/Inference

Computing on Devices

AI

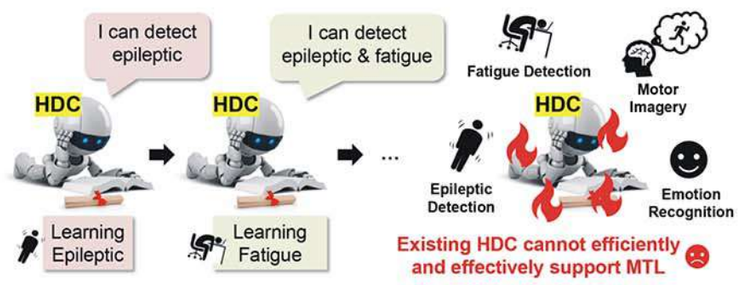
① **Resource constraints on edge devices**
- ❖ Limited computing hardware and memory resources
- ❖ Limited battery capacity
- ❖ Learning is more computation-intensive than inference

➔Model should be 1) low computational complexity and 2) low power in learning and inference

② **Scarcity of labeled data**
- ❖ Only few labeled data from users are available for model personalization and adaption
  - ➢ Siri, google home products, face ID…etc
- ❖ Fit the model with few labeled data

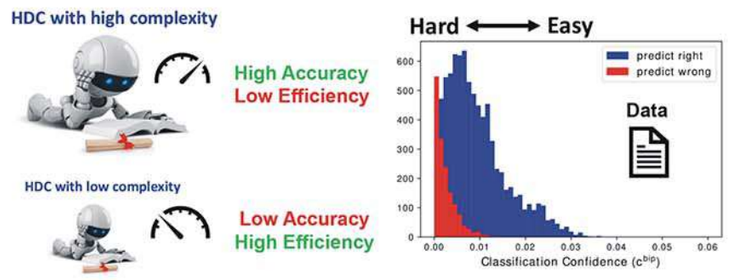
➔ Model should be able to achieve 3) few-shot learning

Issue A. Inefficient to Support Multiple Tasks



- ❖ Due to limited resources available on edge devices, multi-task learning (MTL), which accommodates multiple tasks in one model, is considered a more efficient deployment of HDC
- ❖ However, as the number of tasks increases, MTL-based HDC suffers from the huge memory overhead and performance degradation

Issue B. Accuracy and Efficiency Trade-off

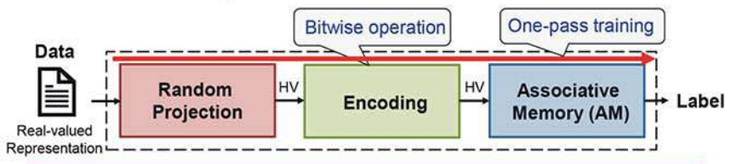


- ❖ Existing **static** HDC limits its trade-off b/w the accuracy and efficiency
  - ❖ Large HDC has high accuracy but low efficiency
  - ❖ Small HDC has high efficiency but low accuracy
- ❖ However, real-world data exhibit varying classification difficulties
  - ❖ Large HDC is unnecessary for easy data ➔ **Potential of dynamic Inference?**

Hyperdimensional Computing (HDC) for Efficient On-Device Computing

- ❖ Proposed by Pentti Kanerva from Berkeley
  - ❖ Brain-inspired computational paradigm
  - ❖ Brain's circuit are massive in the number of neurons
  - ❖ Brain's ability to calculate with binary number is strong

- ❖ Hypervector (HV)
  - ❖ Hyperdimensional vectors (e.g., 10k) ➔ 1 0 1 1 1 … 1
  - ❖ Dense and Binary vectors

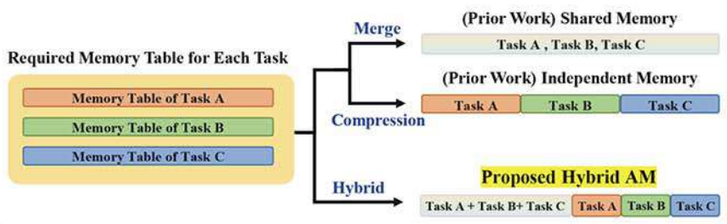


Merits: (1) Efficient computing, (2) Robustness, and (3) Few-shot learning

Proposed MulTa-HDC: A Multi-Task Learning Framework for HDC

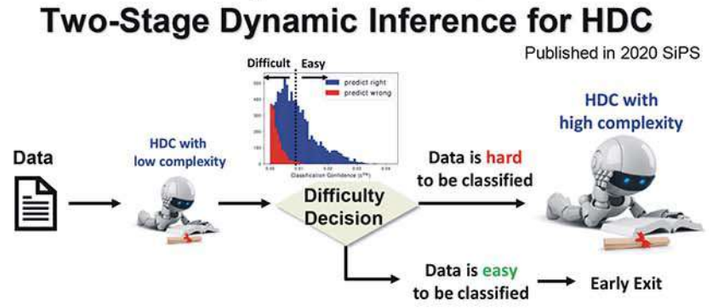
Published in 2020 AIAI, 2020 SiPS, 2021 TC

- ❖ Prior shared memory method causes interference among tasks
- ❖ Prior independent memory method cannot reserve information when memory capacity is insufficient
- ❖ **Proposed MulTa-HDC** leverages the combined strengths of the shared-memory and the independent-memory approaches



Proposed TD-HDC: Two-Stage Dynamic Inference for HDC

Published in 2020 SiPS



- ❖ TD-HDC cascades a low-cost/low-precision HDC, followed by a high-cost/high-precision HDC, to perform **two-stage dynamic inference**
  - ❖ If the low-cost HDC has high confidence in classification results, it allows those easy inputs to exit early
  - ❖ Otherwise, the high-cost HDC is **conditionally** activated for difficult ones

Research Life and Experience

博士研究的道路上充滿著無數漫長且未知的挑戰，有時甚至會在這條道路上迷失方向，但也唯有經歷過這些過程，使我在心態上獲得更多的成長，也讓我更加了解自己的本質與目標。感謝吳安宇指導教授以引導與鼓勵的方式，使我看見研究問題的本質，並以正面的思考方式面對各種困難。感謝父母成為我最強大的後盾，使我能無後顧之憂地專注於研究。感謝Access IC Lab實驗室同伴們，與我一起在研究這條道路上並肩作戰。感謝中技社與評審委員們對我研究的肯定。期許自己在未來的道路上，能善用所學並持續以正向、積極、創新的態度為台灣的科技業以及社會盡一份心力。