Digital & System

PD01

Automatic Alignment of Power Traces for Power Analysis

Yi-Chi Lin[1], Sying-Jyan Wang[1], Chen-Yeh Lin[2], Song-Kong Chong[2], and Katherine Shu-Min Li[3]
[1]Department of Computer Science and Engineering, National Chung Hsing University
[2]Cybersecurity Technology Institute, Institute for Information Industry
[3]Department of Computer Science and Engineering, National Sun Yat-sen University

Data encryption is critical for information security. Previous studies show that power analysis is a powerful tool for side-channel attack against cryptographic modules. Therefore, it is essential to carry out power analysis to assess the vulnerability of cryptographic modules. The success of power analysis relies on aligned power traces, which is not easy without an external trigger point. In this paper, we propose an automatic power trace alignment method so that operations in power traces and be located without external trigger points. Experimental results show that the proposed method can be applied to various ciphers, including AES and RSA.

PD02

Low Complexity and Low Power Sense-Amplifier Flip-Flop for Low Voltage Operation

Jin-Fa Lin, Cheng-Hsueh Yang, Yu-Cheng Yen, Jun-Ting Wu, and Yan-Ying Chen
Department of Information and Communication Engineering, Chaoyang University of Technology

A low power and highly reliable sense-amplifier (SA) based flip-flop (FF) with transition completion detection is proposed. The proposed design integrates the generated detection circuit to indicate the completion of SA stage and thus overcoming the operational yield degradation with 2.3% area saving. Simulation results shown that the minimum VDD of our design is 260mV lower than previous design, which means our design can operate even when VDD is in the subthreshold region.

PD03

Edge-Preserving Filter FPGA Design Based on Side-Window Filter

Jun-Ting Zhang[1], He-Sheng Chou[1], Tsung-Han Lee[1], Chiung-An Chen[2], Szu-Yin Lin[3], Chih-Hsien Hsia[3], and Shih-Lun Chen[1]
[1]Department of Electronic Engineering, Chung Yuan Christian University
[2]Department of Electrical Engineering, Ming Chi University of Technology
[3]Department of Computer Science and Information Engineering, National Ilan University

This paper presents an edge-preserving filter design by using edge detection to reduce the computational complexity of the side window filter. By identifying edge types in advance, the side window filter is unnecessary to produce each result in the same time for comparing. The hardware architecture is implemented on Xilinx's FPGA, which can reduce 86% power consumption, 89% number of LUTs, 67% number of FFs, and 84% number of route nets, respectively. The synthetic results show that this method can reduce the computational complexity of the side window filter and achieve low-cost hardware implementation

PD04

Based on Logarithmic Computing to Design an Embedded CNN Processor

Chong-Yin Lu, Ren-Song Tsay, Wey-shin Chang, and Hsiang-Ting Peng
National Tsing Hua University

This paper proposes an embedded CNN processor design that can easily fit into edge devices based on a modified logarithmic computing method using a deficient bit-width representation. For Yolov2, our processing circuit takes only 0.15mm^2 using TSMC 40 nm cell library. The key idea is to apply a low bit-width logarithmic expression to devise a unified, reusable CNN computing kernel that can significantly reduce computing resources. The proposed approach has been extensively evaluated on many famous image classification CNN models (AlexNet, VGG16 and ResNet-18/34) and object detection models (Yolov4). The hardware-implemented results show that our design achieves 20x performance improvement while consuming only minimal computing and storage resources yet attains very high accuracy. The method is thoroughly verified on FPGAs, and the SoC integration is underway with promising results. Our design is excellent for edge computing purposes with extremely efficient resources and energy usage.

PD05

Smart Crop Growth Monitoring based on System Adaptivity and Edge AI

Chun-Hsian Huang, Shun-Ying Hsieh, Ching-Ting Weng, Min-Fei Hsieh, Shu-Lin Ciou, and Ching-Jui Huang
Department of Computer Science and Information Engineering, National Taitung University

This work proposes a smart crop growth monitoring system that contains an adaptive cryptography engine to ensure the security of sensor data and an edge artificial intelligence (AI) based estimator to classify the pest and disease severity (PDS) of target crops. Based on the smart system management mechanism, cryptographic functions can be adapted to varying and real-time requirements, while the actuators can be controlled to interact with the physical world to ensure the healthy growth of crops. Experiments show when all the four cryptographic hardware modules, including RTEA32, RTEA64, XTEA32 and XTEA64, are supported, using the adaptive cryptography engine, 72.4% of slice LUTs and 68.4% of slice registers in terms of the Xilinx Zynq-7000 XC7Z020 chip can be saved. Furthermore, using the binarized neural network (BNN) hardware module of the PDS estimator, the recognition accuracy of target crops i.e. dragon fruits can achieve 76.57%. Compared to the microprocessor-based design and the GPU accelerated one, the same BNN architecture on the FPGA can accelerate the frames per second by a factor of 4,919.29 and a factor of 1.08, respectively.

PD06

Hybrid Design of Out-of-order Load/Store Instructions of A Risc-V Superscalar

Chun-Wei Chao, Che-Yu Wu, and Chun-Jen Tsai
Department of Computer Science, National Yang Ming Chiao Tung University

In this paper, we present the design of an out-of-order load/store pipeline of the open-source RISC-V superscalar processor, Falco [1]. The original load/store pipeline of Falco only allows in-order execution of memory instructions. However, out-of-order execution of memory instructions is crucial to the performance of superscalars for modern applications. However, the common log-based recovery scheme for speculative mis-prediction usually takes many cycles to recover from a Write-After-Read (WAR) hazard. To improve the load/store pipeline of Falco, we adopt the speculative execution scheme using store sets. In addition, a hybrid recovery mechanism based on the checkpoint snapshots and the log-based instruction rollback is used to reduce the processor recovery time upon mis-prediction. As the experimental results show, the proposed architecture can improve the performance of the original Falco by 18% based on standard benchmarks. The proposed design is verified on a Xilinx FPGA development board and is made open-source for future improvements.

PD07

Low-Complexity Arrhythmia Classification using Phase Portrait of Photoplethysmography with Artificial Neural Network

Po-Lin Yao[1], Shu-Yen Lin[1], and Yu-Wei Chiu[2]
[1]Department of Electrical Engineering and 2Department of Computer Science and Engineering, Yuan Ze University
[2]Cardiology Department, Far Eastern Memorial Hospital

To classify the arrhythmias on the wearable devices, the features from phase portrait of photoplethysmography (PPG) with Artificial Neural Network (ANN) are used because of the low computation complexity and high flexibility. By changing the number of the input features in ANN, the suitable number of layers and neurons in the hidden layers for the wearable devices can be found. A set of 17 input features with 3 hidden layers and the set of 8 features with 1 hidden layer can get up to 97.38% and 86.41% accuracies in bradycardia.

PD08

An Energy Efficiency Architecture Design on Sparsity CNN Accelerator

Chung-Bin Wu and Chung-Hsuan Chen
National Chung-Hsing University

Sparse CNN computing has received attention in recent years, especially for mobile devices, or image detection devices that require high image inference per second. This technology can bring faster processing speed and higher energy efficiency. This paper designs an energy-efficient sparse computing architecture, which can average the workload between each Process Engine (PE) when computing sparse images and can reduce the time for PE to grab valid values when the image is extremely sparse. The proposed hardware can achieve 56.43 GOPS and 4.51 Frame/s under the VGG-16 network at 200Mhz, and the Multiply Accumulate Utilization (MACs Utilization) can reach an average of 97.94%.

PD09

Embedded TCP/IP Controller for a RISC-V SoC

Yi-De Lee, Jung-Chun Tsai, and Chun-Jen Tsai
Department of Computer Science, National Yang Ming Chiao Tung University

In this paper, we present the design of an open-source RISC-V application processor with an embedded TCP/IP network module. Traditionally, the TCP/IP stack is a software layer of the OS kernel due to its complex control behavior. However, previous studies show that a hardwired logic can perform TCP/IP control algorithms much more efficiently than a software implementation. However, to allow a processor to invoke a hardware TCP/IP logic efficiently and create compatible API for existing network applications is not a trivial task. This paper proposes an efficient interface logic between the processor core and the hardware TCP/IP stack through user-defined RISC-V instructions. The proposed architecture is implemented and verified on a Xilinx FPGA development board. Experimental results show that the end-to-end packet delay can be reduced by more than 99% using the proposed network module when compared to a software LWIP stack under a FreeRTOS real-time system. Therefore, the proposed architecture can be very useful for deeply-embedded IOT devices where a low-power processor can be used to handle low-latency high throughput IP packet transmissions.

PD10

Edge AI System for Upper Arm Training

Hsiang-Lung Huang and Ya-Hsin Hsueh
Department of Electronic Engineering, National Yunlin University of Science and Technology

In order to encourage people to exercise in their own homes or private places, this system can help users understand the upper arm training posture correctness of their movements. We build an edge AI motion detection system contains image recognition and a homemade wearable device. In addition to attendant their own action in the real-time on display, users can also see the results of the system's recognition of the movements. When they invention that the correct number of motions has not been accumulated, they can adjust their posture promptly to avoid the incorrect posture for a long time. The combination of image recognition and wearable devices can reduce the misjudgment of movement caused by the environment and other influences. In the era of coronavirus outbreak, this study provides a tool that can help users to check their posture when people choose to exercise at home or in private places without professional guidance.

PD11

Conditional Deep Convolutional GAN for Denoising and Inpainting On-Display Fingerprint

Shi-Xian Zhuang, Chao-Yuan Zheng, and Pei-Yung Hsiao
Department of Electrical Engineering, National University of Kaohsiung

On-Display Fingerprint (ODF) images are usually quite blurry. Nowadays, ODF smartphone has gradually become a standard configuration in the market, the clarity of the fingerprint image obtained by the ODF sensor will seriously affect the accuracy of fingerprint recognition. Therefore, this paper proposes the Deep Convolutional Generative Adversarial Network (DCGAN) based method to effectively improve the quality of ODF images. Moreover, two self-established ODF Ground Truth (GT) databases are used for training and similarity verification. We finally present three DCGAN series network architectures, namely DCGAN, Conditional DCGAN (CDCGAN), and Reconstruction CDCGAN (RecCDCGAN). In addition, the evaluation is based on the SSIM image similarity score and the NFIQ 2 for fingerprint quality estimation. Experiments show that RecCDCGAN can generate the best clear fingerprint images for blurred ODF images.

PD12

Scalable and Reconfigurable Architecture of Modified KD-Tree ML-Classifier with 5-Point Searching

Xin-Yu Shih and Chen-Yen Song
Department of Electrical Engineering, National Sun Yat-sen University

This paper proposes a reconfigurable hardware architecture of modified KD-tree machine-learning classifier. As compared to current literature, this hardware is the first KD-tree-like hardware chip implementation. As compared with original KD-tree algorithm, our design can deliver a very low latency in hardware because we do not need the data traversal steps along the binary tree. Meanwhile, this scalable hardware can be easily constructed if supporting a greater number of data instances to be classified. In the chip implementation with TSMC 40-nm CMOS technology, our reconfigurable hardware chip achieves a maximum frequency of 334.5 MHz, only occupying an area of 0.884 mm2 in APR.

PD13

Latency Minimization for MLP Accelerators using an ILP-Based Weight Allocation Method

Kang-Yi Fan, Jyun-Hua Chen, Wei-Lin Wang, Chien-Nan Liu, and Juinn-Dar Huang
Institute of Electronics, National Yang Ming Chiao Tung University

It is generally impossible to store all weights into an MLP accelerator because of limited on-chip SRAM capacity. However, the performance can still be improved if a portion of weights are allocated in faster SRAM. In this paper, we first present an analytical method for performance evaluation under different weight allocation approaches. We then propose an ILP-based on-chip weight allocation strategy that can maximize the overall performance. Experiment results show that the proposed strategy constantly outperforms several trivial heuristic methods over a large set of various MLP models, MLP accelerator configurations, and on-chip SRAM capacities.

PD14

A FPGA-based System Integration of DPU Unit and Single Image Fog Removal Method Using Improved Dark Channel Prior

Nguyen Hoang Hai Pham, Zhi-Hao Chen, and Chi-Chia Sun
Department of Electrical Engineering, National Formosa University

The current application of deep learning networks in public monitoring systems and driver assistance systems has made great progress. However, it needs a lot of development effort in order to make these neural networks working effectively in a variety of weather. We propose a system integration solution which includes several accelerated digital image processors and Deep Learning Processing Unit (DPU). These digital image processors employ traditional image processing and single image fog removal based on an improved dark channel prior algorithm to enhance input images, while the DPU can run several pruned and quantized deep learning models. We implemented the system on a Field-Programmable Gate Array (FPGA) platform resulting in high accuracy and low power consumption, which is suitable for edge computing or embedded system applications.

PD15

Energy-efficient and Accurate Object Detection Design on FPGA Platform

Kuan-Hung Chen[1], Chun-Wei Su[2], and Jen-He Wang[1]
[1]Department of Electronic Engineering, Feng Chia University
[2]Department of Electronic Engineer, PH.D. program of Electronical and Communications Engineering

With the innovation of hardware equipment, the development of artificial intelligence has broken through the limitations of the past. Neural networks have been continuously deepening to improve the accuracy of detection, so that the parameters have increased with a proportional rate. In this way, however, high power consumption has followed. Therefore, the design of neural network must consider not only detection accuracy but also energy efficiency. In this paper, we analyzed energy consumption, detection accuracy and execution speed of our neural network model as well as the state-of-the-art models based on an FPGA platform called ZCU-102. We adopt the performance index from Low Power Computer Vision (LPCV) challenge which considers power dissipation, mean Average Precision (mAP) and Frames Per Second (FPS) at the same time to evaluate these models in an overall point of view. Agilev4 can achieve 59.9% of mAP@50 on MS COCO test-dev2017 datasets. If the input frame resolution is turned into 416×416, the processing frame rate can reach 20.7 FPS on ZCU-102. Compared with the state-of-the-art models, the LPCV score of Agilev4-416 is 1475.8 which is 1.56 times of that of YOLOv4-416.

PD16

A 1.46TOPS/W Deep Learning Processor with a Reconfigurable Processing Element array based on SRAM Access Optimization

Liao-Chuan Chen, Zhaofang Li, Yi-Jhen, Lin, and Kea-Tiong Tang
National Tsing Hua University

Deep convolutional neural networks feature numerous parameters, causing data movement to usually dominate the power consumed when computing inferences. This paper proposes an on-chip buffer access optimization method and high-data-reuse architecture that can reduce the power consumed by an on-chip buffer by up to 67.8%. The chip is designed in a TSMC 40 nm process running at 200 MHz and achieves energy efficiency of 1.46 TOPS/W.

PD17

MARSv2: Multicore and Programmable Reconstruction Architecture Using SRAM CIM-Based Accelerator with Lightweight Network

Chia-Yu Hsieh[1], Shih-Ting Lin[1], Yen-Wen Chen[1], Chih-Cheng Lu[2], Meng-Fan Chang[1], and Kea-Tiong Tang[1]
[1]Department of Electrical Engineering, National Tsing Hua University
[2]Information and Communication Labs, Industrial Technology Research Institute

Computing-in-memory (CIM) systems reduce the degree of large-scale data movement by performing computation on the memory; this avoids a von Neumann bottleneck. Because of its low-power characteristic, CIM has demonstrated great potential for increasing the energy efficiency of edge devices. This paper presents a multicore and programmable reconstruction architecture using static random-access memory (SRAM) CIM-based accelerator with lightweight network. The proposed architecture uses SRAM CIM macro as the processing element, supporting sparse convolutional neural network computing. This architecture achieves 15.16 TOPS/W system energy efficiency and 747.6 GOPS on the CIFAR10 data set.

PD18

Convolution Neural Network Chip Design using Selective Convolution Layer

Tzu-Huan Huang[1], T. Hui Teo[1], Yan-Lin Lu[2], and I-Chyn Wey[3]
[1]Engineering Product Development, Singapore University of Technology and Design
[2]Electrical Engineering, Chang Gung University
[3]Artificial Intelligence Research Center, School of Electrical and Computer Engineering, College of Engineering, Graduate Institute of Electrical Engineering, Chang Gung University

Engine(SPE-I) is proposed to efficiently accelerate CNN processing by eliminating most unessential operations based on algorithm-hardware coptimizations. First, we will compare two input image similartity. If similartity is too high, this will be regarded as redundant calculations which will be skipped. The experimental results show that accuracy drops by only 0.12%–1.79%with a 73%–81% multiplicative reduction by compared with original CNN model implementations.

PD19

A Winograd Architecture Design for Edge AI Accelerator

Po-Yao Chung and Wei-Kai Cheng
Department of Information and Computer Engineering, Chung Yuan Christian University

With the development of modern artificial intelligence (AI) technology, convolutional neural networks (CNNs) has been widely used in many application domains. However, as the computation demand of CNNs is dominated by convolution layers, some researches exploit Winograd algorithm to mitigate the number of required multiplications. In this paper, we propose a hardware architecture design based on the Winograd algorithm to reduce the computational complexity of the convolutional layer, and we use an approach different from previous Winograd implementations to optimize the time spent in the Winograd matrix conversion process. Except to the stride one CNN computation, our proposed pipeline architecture can also apply to the stride-2 CNN computation based on a decomposition methodology. Compare the efficiency between executing Winograd algorithm on our proposed architecture and executing sliding window convolution on the Systolic array architecture, experiment results show that our architecture can reduce up to 40% of computation cycles under the same number of multiplier–accumulator (MAC) resource.

PD20

Row Echelon Form Reconfigurable Sparse Matrix Elimination Implementation

Bo-Yi Wu
Department of Electrical Engineering, National Cheng Kung University

With the progress of artificial intelligence (AI), the amount of computation data will continue to rise, and more complex calculations will need to be performed on edge devices. More flexible, reconfigurable processing methods, reduced data storage and computation are increasingly important on edge device systems. This paper implements a hardware architecture to find commonalities between filters in convolutional neural network (CNN). Use reconfigurable hardware to support multiple kernel sizes and utilize a systolic array hardware architecture. The Gaussian-Seidel iterative algorithm finds commonality between filters and recombines the filters using linear combination coefficients. Improved Gaussian-Seidel (GS) iteration algorithm to speed up iteration and hardware usage in systolic array hardware architecture. In order to reduce the size of data storage, use the Linked List Encoding (LNK) method to effectively compress filter data of variable kernel size. The design process is based on Algorithm/Architecture co-design, and the data flow is constructed by analyzing the number of operations, hardware parallelism, memory storage, and data transmission. The experimental results show that enhancing the sparsity on operation arrays can reduce the operation on the Resnet50 model. The sparseness of the matrix means that data storage and computation can be reduced through data compression.

PD21

A Novel Fast-Flying Bird Detection and Identification Based on Configurable AI DPU Processor on FPGA Accelerator Card

Afaroj Ahamad, Chi-Chia Sun, and Hoang Minh Nguyen
Digital System Design Laboratory, National Formosa University

In this article, we proposed a deep learning algorithm for fast-flying bird identification methodology and it is implemented on an embedded platform with an Alveo accelerator card. The Alveo accelerator card provides high bandwidth memory architecture. The novel detection method is divided into two sub-processes, i.e. moving object detection and object identification. Accelerated image processing is used to detect moving objects and configurable neural network inference is used for bird identification. The moving object detection process is based on the principle of frame difference. Afterward, the moving object objects are recorded with their size and position within the image. The confirmed moving object is pushed through a deep learning processor unit (DPU) for classification, resulting in the name of the bird species. The proposed method implemented and tested on Alveo U50 accelerated card can attend high accuracy of up to 81 % with the execution of 289 FPS while processing 960x540 resolution videos, and 84 FPS while processing HD definition (1920x1080) videos, furthermore the DPU execution alone can reach 583 FPS.

PD22

The Multi-Phase Direct Digital Synthesizer Based on Pipelined CORDIC Algorithm for Quantum Computer

Tzu-Hsuan Hsu and Hsiao-Chin Chen
National Taiwan University of Science and Technology

This paper presents a direct digital synthesizer (DDS) for the controller in a quantum computer. In order to reduce the hardware cost, the pipeline coordinate rotation digital computer (CORDIC) is adopted, where the hardware cost can be reduced by 45% and the SFDR > 58 dB is achieved. Moreover, according to the simulation, the multi-phase architecture can be used to boost the output frequency range by at least 14 times.

PD23

Deep Learning Accelerator Integration and Implementation of FPGA Platform Handshake Verification

Chung-Bin Wu and Tzu-wei Chan
National Chung-Hsing University

In today's image CNN accelerators, both the number of network layers and the number of operating parameters are complicated to obtain better recognition results. In order to implement DLA (Deep Learning Accelerator) on the Zynq UltraScale+ MPSoC ZCU102 platform, this paper also proposes a related integrated optimization scheme, which is beneficial to the stability of software and hardware.

PD24

Study of Deep-Learning YOLO-Based Driver Monitor System Design with Embedded GPU Devices

Yen-Sok Poon and Chih-Peng Fan
Department of Electrical Engineering, National Chung Hsing University

To develop a non-contact driver behavior detection system to improve driving safety, in this study, by installing a webcam on the dashboard to detect the driver's behavior, and YOLO-based deep learning technology is used. By using RGB channel images as input, YOLO-based deep learning models such as YOLOv3-tiny, YOLOv3-tiny-3l, YOLO-fastest, and YOLO-fastest-xl are adopted and trained as candidate detectors. Detected behaviors include normal driving, distracted head rotation, drowsiness, eating, and telephone conversations. Experimental results show that YOLO-fastest-xl and its lite version can perform best with multi-category datasets when the same parameters are set. By the embedded software implementation on GPU based devices, the proposed design performs 30 frames per second (FPS) for real-time applications.

PD25

Reconfigurable symmetry pattern in Gabor Filter Using Principal Component Analysis with Microprogrammed Controller

Peng-Xiang Wang
Department of Electrical Engineering, National Cheng Kung University

Abstract-With the increasing popularity of artificial intelligence (AI) as AI requires higher precision and large computation, Different data streams can be generated for different algorithms, which can improve efficiency and flexibility this paper is based on algorithm/architecture Co-design (AAC) [2] using computationally efficient Gabor filters via Principal Component Analysis (PCA). PCA projects filter coefficients onto a more symmetric vector space, then reduces the computation by sharing coefficients. On the other hand, in a trade-off between algorithmic accuracy and computational efficiency, we replaced the first convolutional layer of Resnet50. In our experiments, we observed a slight accuracy drop, which is acceptable.