Dongseok Im

dsim@kaist.ac.kr/
LDS1235566@gmail.com

Summary

- **Energy-efficient 2D/3D vision system-on-chip** design experiences
- **Digital circuit-, architecture-, SW-level AI accelerator** design experiences
- **Custom silicon chip and demonstration system board** design experiences
- **Date Available for Work: Mar. 2025**

Work Experience

- KAIST Information & Electronics Research Institute — Sep. 2023 - Present
  Post-doctoral Researcher
  (Technical Research Personnel for Korean Military Duty until Feb. 2025)

Education

- Ph.D. in EE, KAIST (Advisor: Hoi-Jun Yoo) — Mar. 2020 - Aug. 2023
  Thesis Title:
  A Low-Power and Real-Time 3-D Object Recognition System-on-chip
- M.S. in EE, KAIST (Advisor: Hoi-Jun Yoo) — Feb. 2018 - Feb. 2020
- B.S. in EE, POSTECH — Mar. 2014 - Feb. 2018

Awards & Scholarships

- 2022 ISOCC Chip Design Contest Best Poster Award — 2022
- Microsoft Research Asia Fellowship Nomination Award — 2022
- IEEE Solid-State Circuits Society ISSCC Student Travel Grant Award — 2022
- 28th Samsung HumanTech Paper Award (Circuit Design) Bronze Prize — 2022

Activities

- **[Invited Talk]** Samsung AI Forum 2023 Poster Session Presenter 2023
- **[Invited Talk]** Samsung Electronics DRAM Design Team Talk 2023
- **[Invited Talk]** Apple Tech. Talk: "Sibia: Signed Bit-slice Architecture for Dense DNN Acceleration with Slice-level Sparsity Exploitation," Cupertino, CA 2023
- **[Invited Talk]** Apple Tech. Talk: "DSPU: A 281.6mW Real-Time Depth Signal Processing Unit for Deep Learning-Based Dense RGB-D Data Acquisition with Depth Fusion and 3D B-Box Extraction," Cupertino, CA 2022
- **[Journal Reviewer]** IEEE Journal of Solid-State Circuits 2023
- **[Journal Reviewer]** Springer The Journal of Supercomputing 2023
- **[Journal Reviewer]** IEEE Transaction on Circuits and Systems I 2021, 2022
- **[System Demo]** ACM/IEEE ISLPED Design Contest Presenter 2022
- **[System Demo]** IEEE ISSCC Demonstration Session Presenter 2022
- **[Student Chair]** Student chair of the KAIST Semiconductor System Lab 2021-2022
- **[International Exchange Program]** Summer Session, UC Berkeley 2016

Skills & Languages

- Verilog, C/C++, Python, MATLAB, Java (Android API), C# (Unity API)
- Tensorflow, Pytorch
- Semi-custom Design Flow in Synopsys
- FPGA (Altera/Xilinx) Design Tool Chains
- Schematic Design in OrCAD Capture
- SoC Platform Design with RISC-V ISA
- Native Korean / Fluent English

Research Experience

Hardware/Software Development for Deep Learning-based 3D Vision System

1. Designed a silicon chip for a real-time and low-power **end-to-end 3D Vision SoC** (Representative Paper: ISSCC 2022, 1st Author)
   - (SW) Low-power monocular depth estimation + low-resolution ToF sensor fusion + 3D object detection system

- (SW/HW) Flexible point processing unit with the flexible window search-based algorithm for efficient 3D point sampling and grouping
- (SW/HW) Flexible matrix processing unit with the bit-slice-level dual sparsity exploitation for efficient 2D/3D DNN and GEMM accelerations
- (SW/HW) 3D data correlation matrix codec unit with the hierarchical encoding and decoding for minimized data transactions

2. Designed a silicon chip for an energy-efficient **3D object detection processor** (Representative Paper: S.VLSI 2021)
   - (SW/HW) Page-based sparse 3D point data management unit with the block-wise point processing algorithm
3. Designed a silicon chip for a low latency **3D hand pose estimation processor** (Representative Paper: S.VLSI 2020, 1st Author)
   - (SW/HW) Pipelined 3D DNN architecture with the large-scale max pooling speculation

Hardware/Software Development for Deep Learning-based 2D Vision System

1. Designed a silicon chip for an energy-efficient **fixed-point-based DNN training processor** (Representative Paper: JSSC 2021)
   - (SW) Stochastic dynamic fixed-point and bit-precision searching methods
2. Designed a silicon chip for **a generative adversary network training processor** (Representative Paper: ISSCC 2020)
   - (HW) Programmable scale-out architecture for multi-DNN acceleration with computation-bandwidth optimization
   - (HW) DNN core with dual sparsity exploitation for inference & training
3. Developed a hardware architecture for **an efficient image segmentation processor** (Representative Paper: TCAS-I 2020, 1st Author)
   - (HW) Delay cell-based systolic array for dilated and transposed convolution
   - (SW) Dilation rate adjustment algorithm on dilated convolution by a size of image region-of-interest for low-latency image segmentation

Energy-efficient Bit-scalable Hardware Architecture Designs

1. Designed a dense-sparse DNN architecture with the **Radix-4 LUT-based computation** (Representative Paper: HPCA 2024, 1st Author)
2. Designed a dense DNN architecture with the **lossless signed bit-slice representation** (Representative Paper: HPCA 2023, 1st Author)
3. Designed a multi-DNN architecture with the **bit-slice-level dual sparsity exploitation** (Representative Paper: ISSCC 2022, 1st Author)

4. Designed a multi-precision DNN training architecture with the **dynamic bit-precision searching and bit-slice-level sparsity exploitation**
   (Representative Paper: JSSC 2021)

Instruction Set Architecture (ISA) Compatible SoC Designs

1. Designed a silicon chip for **a 3D vision SoC with RISC-V ISA**
   (Representative Paper: ISSCC 2022, 1st Author)
2. Designed a silicon chip for a **2D image super resolution SoC with RISC-V ISA**
   (Representative Paper: CICC 2022)

Demonstration System Board Design

1. Designed a **3D object detection demonstration system** using ASIC and Android API
   - Developed a custom C/C++ and Java libraries with HW system (DNN ASIC, FPGA, HyperBUS Memory, 2D/3D Image Sensors, and USB controller)
2. Designed a **3D hand pose estimation demonstration system** using ASIC and Unity API
   - Developed a custom C/C++ and C# libraries with HW system (DNN ASIC, FPGA, DDR3, and USB controller)

Publications

International Journals

1. **[Micro]** D. Im, G. Park, Z. Li, J. Ryu, S. Kang, D. Han, J. Lee, W. Park, H. Kwon, HJ. Yoo
   A Mobile 3D Object Recognition Processor with Deep Learning-based Monocular Depth Estimation, IEEE Micro, 2023 (Invited)
2. **[JSSC]** Z. Li, S. Kim, D. Im, D. Han, HJ. Yoo
   An Efficient Deep-Learning-Based Super-Resolution Accelerating SoC With Heterogeneous Accelerating and Hierarchical Cache, IEEE Journal of Solid-State Circuits, 2022
3. **[JSSC]** D. Im, G. Park, J. Ryu, Z. Li, S. Kang, D. Han, J. Lee, W. Park, H. Kwon, HJ. Yoo
   DSPU: An Efficient Deep Learning-Based Dense RGB-D Data Acquisition with Sensor Fusion and 3-D Perception SoC, IEEE Journal of Solid-State Circuits, 2022 (Invited)
4. **[Micro]** D. Han, D. Im, G. Park, Y. Kim, S. Song, J. Lee, HJ. Yoo
   A Mobile DNN Training Processor with Automatic Bit-precision Search and Fine-grained Sparsity Exploitation, IEEE Micro, 2021
5. **[JSSC]** D. Im, D. Han, S. Kang, HJ. Yoo
   A Pipelined Point Cloud Based Neural Network Processor for 3-D Vision With Large-Scale

Max Pooling Layer Prediction, IEEE Journal of Solid-State Circuits, 2021

6. **[JSSC]** S. Kang, D. Han, J. Lee, D. Im, S. Kim, S. Kim, J. Ryu, HJ. Yoo
   GANPU: An energy-efficient multi-DNN training processor for GANs with speculative dual-sparsity exploitation, IEEE Journal of Solid-State Circuits, 2021
7. **[JSSC]** D. Han, D. Im, G. Park, Y. Kim, S. Song, J. Lee, HJ. Yoo
   An Energy-Efficient Deep Neural Network Training Processor with Bit-Slice-Level Reconfigurability and Sparsity Exploitation, IEEE Journal of Solid-State Circuits, 2021
8. **[TCAS-II]** J. Ryu, G. Park, D. Im, JH. Kim, HJ. Yoo
   A 0.82 µW CIS-Based Action Recognition SoC With Self-Adjustable Frame Resolution for Always-on IoT Devices, IEEE Transactions on Circuits and Systems II, 2021
9. **[TCAS-II]** G. Park, D. Im, D. Han, HJ. Yoo
   A 1.15 TOPS/W energy-efficient capsule network accelerator for real-time 3D point cloud segmentation in mobile environment, IEEE Transactions on Circuits and Systems II, 2020
10. **[TCAS-I]** D. Im, D. Han, S. Choi, S. Kang, HJ. Yoo
    DT-CNN: An energy-efficient dilated and transposed convolutional neural network processor for region of interest based image segmentation, IEEE Transactions on Circuits and Systems I, 2020

International Conferences

1. **[HPCA]** D. Im, HJ. Yoo
   LUTein: Dense-Sparse Bit-slice Architecture with Radix-4 LUT-based Slice-Tensor Processing Units, IEEE International Symposium on High-Performance Computer Architecture, 2024
2. **[ISSCC]** G. Park, S. Song, H. Sang, D. Im, D. Han, S. Kim, H. Lee, HJ. Yoo
   Space-Mate: A 303.5mW Real-Time Sparse Mixture-of-Experts-Based NeRF-SLAM Processor for Mobile Spatial Computing, IEEE International Conference on Solid-State Circuits, 2024
3. **[ISSCC]** J. Ryu, H. Kwon, W. Park, Z. Li, B. Kwon, D. Han. D. Im, S. Kim, H. Joo, HJ. Yoo
   NeuGPU: A 18.5 mJ/iter Neural-Graphics Processing Unit for Instant-Modeling and Real-Time-Rendering with Segmented-Hashing Architecture, IEEE International Conference on Solid-State Circuits, 2024
4. **[ASSCC]** J. Park, D. Han, J. Ryu, D. Im, G. Park, HJ. Yoo
   A 33.58 FPS Embedding based Real-time Neural Rendering Accelerator with Switchable Computation Skipping Architecture on Edge Device, IEEE Asian Solid-State Circuits Conference, 2023
5. **[S.VLSI]** W. Xie, H. Sang, B. Kwon, D. Im, S. Kim, S. Kim, HJ. Yoo
   A 709.3 TOPS/W Event-Driven Smart Vision SoC with High-Linearity and Reconfigurable MRAM PIM, IEEE Symposium on VLSI Technology and Circuits, 2023

6. **[HPCA]** D. Im, G. Park, Z. Li, J. Ryu, HJ. Yoo
   Sibia: Signed Bit-slice Architecture for Dense DNN Acceleration with Slice-level Sparsity Exploitation, IEEE International Symposium on High-Performance Computer Architecture, 2023
7. **[ASSCC]** W. Park, D. Im, H. Kwon, HJ. Yoo
   An Efficient Unsupervised Learning-based Monocular Depth Estimation Processor with Partial-Switchable Systolic Array Architecture in Edge Devices, IEEE Asian Solid-State Circuits Conference, 2022
8. **[HOTCHIPS]** Z. Li, S. Kim, D. Im, D. Han, HJ. Yoo
   An Energy-efficient High-quality FHD Super-resolution Mobile Accelerator SoC with Hybrid-precision and Energy-efficient Cache Subsystem, IEEE Hot Chips Symposium, 2022
9. **[HOTCHIPS]** D. Im, G. Park, Z. Li, J. Ryu, S. Kang, D. Han, J. Lee, W. Park, H. Kwon, HJ. Yoo
   DSPU: A 281.6mW Real-Time Deep Learning-Based Dense RGB-D Data Acquisition with Sensor Fusion and 3D Perception System-on-Chip, IEEE Hot Chips Symposium, 2022
10. [**HOTCHIPS**] D. Han, D. Im, G. Park, Y. Kim, S. Song, J. Lee, HJ. Yoo
    HNPU-V2: A 46.6 FPS DNN Training Processor for Real-World Environmental Adaptation based Robust Object Detection on Mobile Devices, IEEE Hot Chips Symposium, 2022
11. [**AICAS**] D. Han, D. Im, G. Park, Y. Kim, S. Song, J. Lee, HJ. Yoo
    A 0.95 mJ/frame DNN Training Processor for Robust Object Detection with Real-World Environmental Adaptation, IEEE International Conference on Artificial Intelligence Circuits and Systems, 2022
12. **[COOLCHIPS]** D. Im, G. Park, J. Ryu, Z. Li, S. Kang, D. Han, J. Lee, W. Park, H. Kwon, HJ. Yoo
    A Low-power and Real-time 3D Object Recognition Processor with Dense RGB-D Data Acquisition in Mobile Platforms, IEEE Symposium on Low-Power and High-Speed Chips and Systems, 2022
13. **[CICC]** Z. Li, S. Kim, D. Im, D. Han, HJ. Yoo
    An 0.92 mJ/frame High-quality FHD Super-resolution Mobile Accelerator SoC with Hybrid-precision and Energy-efficient Cache, IEEE Custom Integrated Circuits Conference, 2022
14. **[ISSCC]** D. Im, G. Park, Z. Li, J. Ryu, S. Kang, D. Han, J. Lee, HJ. Yoo
    DSPU: A 281.6mW Real-Time Depth Signal Processing Unit for Deep Learning-Based Dense RGB-D Data Acquisition with Depth Fusion and 3D Bounding Box Extraction in Mobile Platforms, IEEE International Conference on Solid-State Circuits, 2022
15. **[IEDM]** J. Ryu, D. Im, HJ. Yoo
    AI SoCs for AR/VR User-Interaction, IEEE International Electron Devices Meeting, 2021
16. **[HOTCHIPS]** S. Kim, J. Lee, D. Im, HJ. Yoo
    PNNPU: A Fast and Efficient 3D Point Cloud-based Neural Network Processor with Block-

based Point Processing for Regular DRAM Access, IEEE Hot Chips Symposium, 2021

17. **[S.VLSI]** S. Kim, J. Lee, D. Im, HJ. Yoo
    PNNPU: A 11.9 TOPS/W High-speed 3D Point Cloud-based Neural Network Processor with Block-based Point Processing for Regular DRAM Access, IEEE Symposium on VLSI Technology and Circuits, 2021
18. **[ISCAS]** J. Ryu, G. Park, D. Im, JH. Kim, HJ. Yoo
    A 0.82 µW CIS-Based Action Recognition SoC With Self-Adjustable Frame Resolution for Always-on IoT Devices, IEEE International Symposium on Circuits and Systems, 2021
19. **[ISCAS]** Z. Li, D. Im, J. Lee, HJ. Yoo
    A 3.6 TOPS/W Hybrid FP-FXP Deep Learning Processor with Outlier Compensation for Image-to-image Application, IEEE International Symposium on Circuits and Systems, 2021
20. **[COOLCHIPS]** D. Han, D. Im, G. Park, Y. Kim, S. Song, J. Lee, HJ. Yoo
    An Energy-Efficient Deep Neural Network Training Processor with Bit-Slice-Level Reconfigurability and Sparsity Exploitation, IEEE Symposium on Low-Power and High-Speed Chips and Systems, 2021
21. **[HOTCHIPS]** S. Kang, D. Han, J. Lee, D. Im, S. Kim, S. Kim, J. Ryu, HJ. Yoo
    GANPU: A Versatile Many-Core Processor for Training GAN on Mobile Devices with Speculative Dual-Sparsity Exploitation, IEEE Hot Chips Symposium, 2020
22. **[S.VLSI]** D. Im, S. Kang, D. Han, S. Choi, HJ. Yoo
    A 4.45 ms Low-latency 3D Point-cloud-based Neural Network Processor for Hand Pose Estimation in Immersive Wearable Devices, IEEE Symposium on VLSI Technology and Circuits, 2020
23. **[ISSCC]** S. Kang, D. Han, J. Lee, D. Im, S. Kim, S. Kim, HJ. Yoo
    GANPU: A 135TFLOPS/W multi-DNN training processor for GANs with speculative dual-sparsity exploitation, IEEE International Conference on Solid-State Circuits, 2020
24. **[ISCAS]** D. Im, D. Han, S. Choi, S. Kang, HJ. Yoo
    DT-CNN: Dilated and transposed convolution neural network accelerator for real-time image segmentation on mobile devices, IEEE International Symposium on Circuits and Systems, 2019