(858) 405-2529
San Diego, CA
janarbek@gmail.com

Janarbek Matai

Machine Learning Researcher/Engineer

Website: janarbek.github.io/
https://www.linkedin.com/in/

- Strong ML model development model performance optimization expertise balancing trade-off computational efficiency
- Expertise in machine learning workloads such as Vision transformers, LLMs and LVMs and their optimizations.
- Expertise in both ML model development and optimizing performance of models in various hardware targets such as GPU/NPU.
- Published, patented in top tier conferences (Computer vision: ICCV, CVPR, BMBC, Neurips, HW conferences: FPGA, FCCM, DAC,)
- Can work as an independent contributor developing system designs for ML, architecting HW systems for ML, or managing the super engineers/researchers to get challenging projects completed.
- Provided technical leadership roles in my current and previous positions and educated senior leaders and mentored junior engineers, brought multiple academic collaborations and influenced hiring top talent

Education

Ph.D. in Computer Science, *University of California, San Diego* — Mar 2015
M.S. in Computer Science, *Korea Advanced Institute of Science and Technology (ICU)* — Feb 2007
B.S. in Computer Science, *Mongolian University of Science and Technology* — Feb 2004

Skills

Programming	Python, C, C++
Machine Learning Tools	PyTorch, Weight and Bias, ML Compilers
Tools	Git, Continuous Integration, LLVM (Limited exposure), OpenCV, MATLAB, Latex, Valgrind, Clang
Parallel Programming	OpenCL (Limited exposure), CUDA
Hardware	High-Level Synthesis (HLS), VHDL/Verilog(limited exposure), FPGA design, Xilinx/Altera tools, Modelsim
Misc	Embedded Linux Development with Yocto, Working knowledge of PCIe, AXI, SPI, IIC

Experiences

Principal Member of Technical Staff — **July 2024 — Present**
Advanced Micro Devices, — *San Diego, CA*

- Build ML models that push state-of-the-art either using proprietary computer vision/video dataset (video frame interpolation)
- Design and optimization of model performance and accuracy for computational trade-off visual applications (neural texture compression)
- Lead an research effort for using Transformers and State-Space Models for small and large datasets

Sr. Staff Engineer — **Oct 2019 — June, 2024**
Qualcomm AI Research A — *San Diego, CA*

- Delivered research on HW-SW co-design of deep learning for low-power devices and camera ISP. This project leads to starting of an R&D effort that facilitates HW-SW co-design project for machine learning and computer vision across departments.
- Delivered and optimized CNNs and Transformers using various techniques ranging from quantization to algorithmic efficiency
- Optimizing architectural trade-off between hardware and algorithm co-design
- Initiated and lead an R&D for the design and implementation of 3D computer vision on edge devices. Successfully delivered Neurips 2021 demo.
- Established R&D collaboration effort between Qualcomm AI with Universities (3D with UCSD, HW-SW co-design with Cornell). These collaborations resulted in multiple successful publications /demos in CVPR/ BMVC/ NeurIPS , and resulted in the hiring of top-performing interns.
- Leading an R&D effort that facilitates machine learning compiler design for embedded systems
- Delivered design and implementation of machine learning compiler pass for conditional compute

Lecturer (Held in quarterly basis) — **Sep 2015 — Present**
Department of Computer Science and Engineering, University of California, San Diego — *San Diego, CA*

- Lecturing for FPGA design with High-Level Synthesis for signal processing applications
- Lecturing for software for embedded systems class for signal processing applications
- Developed labs and lead the discussion for embedded systems labs

Principal Software Engineer — **Feb 2018 — Aug 2019**
Cognex Corporation, Advanced R&D Lab — *San Diego, CA*

- Design and implementation of high-performance (quantized) neural network on an FPGA
- Design and implementation of high-performance computer vision systems
- Leading an R&D effort that facilitates collaboration between University and Cognex.
- Design of SW/HW co-design systems for vision algorithms.

(858) 405-2529
San Diego, CA
janarbek@gmail.com

Janarbek Matai

Machine Learning Researcher/Engineer

Website: janarbek.github.io/
https://www.linkedin.com/in/

Senior Software Engineer **Mar 2015 — Feb 2018**
Cognex Corporation, Advanced R&D Lab *San Diego, CA*

- Design and implementation of neural network algorithm on an FPGA
- Designing FPGA/GPU systems for high-speed cameras.

Assistant Adjunct Professor **Mar 2017 — Oct 2017**
Department of Computer Science and Engineering, University of California, San Diego *San Diego, CA*

- Lecturing graduate level courses in signal and image processing applications

Research Intern **Jun 2013 — Sep 2013**
Microsoft Research *Redmond, WA*

- Designed canonical Huffman encoding on an FPGA

Research Intern **Jul 2011 — Oct 2011**
Xilinx Research Lab *Dublin, Ireland*

- Designed Viola and Jones based face detection system on an FPGA

Researcher **Feb 2007 — Apr 2009**
Electronics and Telecommunications Research Institute *Daejon, S. Korea*

- Research and development focusing on networked robotics.

Publications

Theses:

1. **J. Matai** , "Templates and Patterns: Augmenting High-Level Synthesis for Domain-Specific Computing," *PhD Thesis, Department of Computer Science and Engineering, University of California, San Diego*, March 2015.

Books:

1. R. Kastner, **J. Matai** S. Neuendorffer, "Parallel Programming for FPGAs," *http://hls.ucsd.edu/*

Journals:

1. A. Irturk, **J. Matai**, J. Oberg, J. Su, R. Kastner, "Simulate and Eliminate: A Top-to-Bottom Design Methodology for Automatic Generation of Application Specific Architectures,"*IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 30, issue 8, August 2011

2. HM. Do, **J. Matai**, YH. Suh, YS. Kim, BK. Kim, HS. Kim, T. Tanikawa, K. Ohba, JY. Lee and W. Yu, "Connection Framework of RT-Middleware and CAMUS for Maintaining Ubiquity between Two Ubiquitous Robot Spaces,"*Advanced Robotics*, vol. 23, issue 12, 2009.

Peer-Reviewed Conference and Workshop Publications:

1. L. Wu, R. Zhu, M. Yaldiz, Y. Zhu, H. Cai, **J. Matai**, F. Porikli, T. Li, M. Chandraker, R. Ramamoorthi ," Factorized Inverse Path Tracing for Efficient and Accurate Material-Lighting Estimation," *International Conference on Computer Vision, ICCV 2023*

2. R. Huang, Z. Yue, C. Huang, **J. Matai**, Z. Zhang," Performance Analysis of Binary Neural Networks Deployed in NVM Crossbar Architectures," *Fourth Workshop on Benchmarking Machine Learning Workloads on Emerging Hardware, MLsys 2023*

3. S Kinzer, S Ghodrati, R Mahapatra, BH Ahn, E Mascarenhas, X Li, **J Matai**, L. Zhang, H. Esmaeilzadeh," Restoring the Broken Covenant Between Compilers and Deep Learning Accelerators," *Archive 2023*

4. R. Zhu, Z. Li, **J. Matai**, F. Porikli, M. Chandraker" IRISformer: Dense Vision Transformers for Single-Image Inverse Rendering in Indoor Scenes," *Conference on Computer Vision and Pattern Recognition (CVPR 2022)*

5. C. Hong, **J. Matai**, S. Borse, Y. Zhang, A. Ansari, F.Porikli, " X-Distill: Improving Self-Supervised Monocular Depth via Cross-Task Distillation," *The British Machine Vision Conference (BMVC)*, Nov 2021

6. **J. Matai**, D. Richmond, D. Lee, Z. Blair, Q.Wu, A. Abazari and R. Kastner, " Resolve: Computer Generation of High-Performance Sorting Architectures from High-Level Synthesis," *International Symposium on Field Programmable Gate Arrays (FPGA)*, February 2016 - **Acceptance Rate 20/105 = 19%**

7. **J. Matai**, D. Lee, A. Althoff and R. Kastner, "Composable, Parameterizable Templates for High Level Synthesis," *Design Automation and Test in Europe (DATE)*, March 2016 - **Acceptance Rate 199/829 = 24%**

8. B. Mao, W. Hu, A. Althoff, **J. Matai**, J. Valamehr, T. Sherwood, D. Mu, and R.Kastner, "Quantifying Timing-Based Information Flow in Cryptographic Hardware," *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)-accepted*

9. Q. Gautier, A.Shearer, **J. Matai**, D. Richmond, P. Meng and R.Kastner, "Real-time 3D Reconstruction for FPGAs: A Case Study for Evaluating the Performance, Area, and Programmability Trade Offs of the Altera OpenCL SDK," *International Conference on Field-Programmable Technology*, December 2014

10. **J. Matai**, D. Richmond, D. Lee and R. Kastner, "Enabling FPGAs for the Masses," *First International Workshop on FPGAs for Software Programmers*, September 2014.

11. D. Lee, **J. Matai**, B. Weals and R. Kastner, "High Throughput Channel Tracking for JTRS Wireless Channel Emulation," 24th *International Conference on Field Programmable Logic and Applications*, September 2014

12. **J. Matai**, JY. Kim and R. Kastner, "Energy Efficient Canonical Huffman Encoding," 25th *IEEE International Conference on Application-specific Systems, Architectures and Processors*, June 2014 - **Acceptance Rate 22/85 = 25.9%.**

13. M. Kimura, **J. Matai**, M. Jacobsen and R. Kastner, "A Low-Power AdaBoost-Based Object Detection Processor Using Haar-Like Features," *IEEE International Conference on Consumer Electronics*, September 2013.

14. **J. Matai**, P. Meng, L. Wu, B. Weals and R. Kastner, "Designing a Hardware in the Loop Wireless Digital Channel Emulator for Software Defined Radio," 11th *International Conference on Field-Programmable Technology*, December 2012 - **Acceptance Rate 24/114 = 21%**

15. **J. Matai**, J. Oberg, A. Irturk, T. Kim and R. Kastner, "Trimmed VLIW: Moving Application Specific Processors Towards High Level Synthesis," *The Electronic System Level Synthesis Conference*, June 2012.

16. **J. Matai**, A. Irturk and R. Kastner, "Design and Implementation of an FPGA-based Real-Time Face Recognition System," *IEEE* 19th *Annual International Symposium on Field-Programmable Custom Computing Machines*, May 2011 - **Acceptance Rate: 42/119 = 35.3%**

17. HM. Do, **J. Matai**, YH. Suh, YS. Kim, BK. Kim, HS. Kim, T. Tanikawa, K. Ohba, JY. Lee and W. Yu, "Connection methodology for two ubiquitous robot spaces - connection of RT-Middleware and CAMUS," *IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, July 2008.

18. DS. Han, JS. Song, **J. Matai** and MK. Lee, "Stress Prediction System for Mobile U-health," 9th *International Conference on e-Health Networking, Applications and Services*, june 2007.

19. **J. Matai** and DS. Han, "Learning-Based Trust Model for Optimization of Selecting Web Services," 9th *Asia-Pacific Web Conference*, May 2007.

Activities

1. Reviews

   - Journal reviewer: Embedded Systems Letters, International Journal of Reconfigurable Computing
   - External reviewer: ICCD 2011, FPL 2011, FPL 2013, FPL 2014, ASAP 2014

2. Invited Participants

   - Amazon Research Symposium, Seattle, WA 2014
   - Latin American eScience Workshop, Sao Paulo, Brazil 2013
   - Astana start-up weekend, Astana, Kazakhstan, 2012