State of AI and ML-Spring 2019

In this event as part of series of regularly planned events, we plan to cover the state-of-the art advances in AI technology. We focus on AI Accelerators, Self-Driving and Face Processing. We feature five thought leaders from Computing, Autonomous systems and Computer Vision.

April 4th 2019, Thursday, 1:00 PM – 7:00 PM (Pacific Time)

SC12 Auditorium, 3600 Juliette Ln,

Santa Clara, CA 95054

Each talk is 45 minutes followed by 10 minutes for Q&A.

Check In: 1.00pm-1.30pm

Key Note: "Emergence of RISC-V computer architecture: embedded applications and machine learning acceleration",
Dr. Zvonimir Bandić, Sr. Director of Next Generation Platform Technologies at Western Digital Corporation, Board of Directors member of RISC-V foundation, Chair of OpenCAPI foundation. 1.30pm-2.25pm

Talk 2: "Accelerating Deep Neural Network Inference with FPGAs",
Mr. Rahul Nimaiyar, Director, Data Center IP Solutions at Xilinx. 2.30pm-3.25pm

Talk 3: "An ASIC approach to unlock deep learning innovation",
Dr. Carlos Macian, Sr. Director AI Strategy and Products at eSilicon. 3.30pm-4.25pm

Break for Networking and Snacks: 4.25pm-5.00pm

Talk 4: "Expectations from an AI Co-Processor for ADAS and Autonomous Driving",
Dr. Vikram Narayan, Head of AI & Computer Vision, the ADAS Group of Visteon Corporation, 5.00pm-5.55pm

Talk 5: "Introduction to Face Processing with Computer Vision",
Mr. Gabriel Bianconi, Founder, Scalar Research, 6:00pm-6.55pm

Program Chair: Dr. Kiran Gunnam, Distinguished Engineer - Machine Learning & Computer Vision

ValleyML.ai State of AI and ML-Spring 2019

Visit ValleyML.ai for other related events and hands-on workshops.

Key Note: "Emergence of RISC-V computer architecture: embedded applications and machine learning acceleration"

Dr. Zvonimir Z. Bandić, Sr. Director of Next Generation Platform Technologies at Western Digital Corporation, Board of Directors member of RISC-V foundation, Chair of OpenCAPI foundation

Abstract: RISC-V Instruction Set Architecture has recently become a key driver for driving open source projects across wide gamut of end applications. Most recently we have seen a lot of application in the Internet of Things (IoT) segment, microcontrollers for a variety of traditional embedded applications, and applications requiring capability for low power operation of inference engines based on artificial neural networks.

We have developed a super-scalar (2-way), 9-stage pipeline, mostly in-order, open-source core based on the RISC-V RV32IMC instructions set, named SweRV. It initially targets in-house embedded Storage System on Chip applications. We present some of the architectural details of the core and implementation challenges, as well as discuss application of the core for the Flash controllers. We also report performance measurements of Coremark and Dhrystone benchmarks, which are traditionally used for embedded core performance benchmarking. Some of the implementation challenges that we have encountered were related to the tradeoff of code density and performance of RISC-V. We report our initial findings and code density improvement solutions based on compiler and linker optimizations.

We are witnessing computation shifting to dedicated machine learning and inference accelerators, typically attached to high speed peripherals buses (PCI express, OpenCAPI) or smart networking protocols (OmniXtend). These systems are typically implemented by high density arrays of multiple and accumulate (MAC) elements. Some of implementations are based on RISC-V Cores implementing vector instruction set extensions. We will review vector set instructions in RISC-V and highlight the architectural value for machine learning and neural network inferencing workloads.

Bio: Zvonimir Z. Bandić is a research staff member and senior director of Next Generation Platform Technologies at Western Digital Corporation in San Jose, Calif. He received his Bachelor of Science in electrical engineering in 1994 from the University of Belgrade, Yugoslavia, and his Master of Science (1995) and PhD (1999) in applied physics from Caltech, Pasadena, in the field of novel electronic devices based on wide bandgap semiconductors. He is currently focusing on both NAND and emerging Non-Volatile Memories (PCM, ReRAM, MRAM) applications for data center storage and computing, including CPU, memory, networking and storage. He has been awarded over 50 patents in the fields of solid-state electronics, solid state disk controller technology, security architecture and storage systems and has published over 50 peer-reviewed papers. He is also Board of Directors member of RISC-V foundation.

Western Digital creates environments for data to thrive. The company is driving the innovation needed to help customers capture, preserve, access and transform an ever-increasing diversity of data. Everywhere data lives, from advanced data centers to mobile sensors to personal devices, our industry-leading solutions deliver the possibilities of data. Western Digital® data-centric solutions are marketed under the G-Technology™, HGST, SanDisk®, Tegile™, Upthere™, and WD® brands.

ValleyML.ai State of AI and ML-Spring 2019

Visit ValleyML.ai for other related events and hands-on workshops.

Talk 2: "Accelerating Deep Neural Network Inference with FPGAs"

Mr. Rahul Nimaiyar, Director, Data Center IP Solutions at Xilinx

Abstract: This talk will cover Deep Learning inference optimizations for wide scale deployment. We will describe both the algorithm techniques as well as hardware specialization for accelerating Deep Neural Network.

Bio: Rahul Nimaiyar is Director, Data Center IP Solutions at Xilinx. He leads development of Machine Learning solutions including Deep Learning Inference and traditional Machine Learning Inference for Cloud and Data Center. Rahul received his BSEE from Indian Institute of Technology, Kharagpur and earned an MBA from Kellogg School of Management at Northwestern University.

We Are Building the Adaptable, Intelligent World. Xilinx is the inventor of the FPGA, programmable SoCs, and now, the ACAP. Our highly-flexible programmable silicon, enabled by a suite of advanced software and tools, drives rapid innovation across a wide span of industries and technologies - from consumer to cars to the cloud. Xilinx delivers the most dynamic processing technology in the industry, enabling rapid innovation with its adaptable, intelligent computing.

Talk 3: "An ASIC approach to unlock deep learning innovation"

Dr. Carlos Macian, Senior Director AI Strategy and Products at eSilicon

Abstract: AI/deep learning algorithms demand superior performance. While FPGAs, GPUs and other forms of dedicated processor help, a custom chip always provides the ultimate in performance with the lowest power and area. The challenge is mapping advanced and rapidly evolving algorithms to an ASIC in a predictable and cost-effective manner. eSilicon will present a unique approach to address this challenge. Ever wonder how Facebook's facial recognition or Snapchat's filters work? In this talk, we'll help you understand some of the computer vision and machine learning techniques behind these applications, and use this knowledge to develop our own prototypes for the products above.

Bio: Carlos Macian is Senior Director AI Strategy and Products at eSilicon. Carlos Macian advises eSilicon on new technologies and innovation, with an emphasis on the AI product space. AI is a rapidly evolving field. Through eSilicon's neuASIC™ IP platform, we are providing the IP and architecture necessary to track changing AI algorithms to support more efficient AI accelerator development through ASIC technology. He previously was the senior director of innovation and the director of operations for eSilicon's Europe and the Middle East organizations. Prior to eSilicon, Dr. Macian was an assistant professor at Universitat Pompeu Fabra, Barcelona, Spain, and an R&D engineering manager for Elastix Corporation. Dr. Macian's career began as an R&D engineer at the Institute of Communication Networks and Computer Engineering in Stuttgart, Germany. Dr. Macian advises technology startups in the Barcelona area. He holds a BS in International Relations from the London School of Economics and Political Science, a BS and MS/EE from the Universitat Politècnica de València and a Ph.D. in computer networks from the University of Stuttgart. He's also a certified professional basketball coach.

Talk 4: "Expectations from an AI Co-Processor for ADAS and Autonomous Driving"

Dr. Vikram Narayan, Head of the AI & Computer Vision Team at the ADAS Group of Visteon Corporation

Abstract: The onset of deep learning approach to solve classification problems are being applied in the automotive space, especially for camera input. In parallel, the camera sensor resolution is increasing dramatically offering higher detection range. Given this a new family of SoCs or the so called

"AI Co-Processors" are being explored by many SoC vendors, Tier-1s and technology companies. Despite the academic appeal, there is a lot to bridge between the proof-of-concept to actual production system. In this talk, we will explore and analyze many practical issues that the engineering encounters while working on developing AI based SW stack on a specific AI Co-Processor. I will dwell into topics like: (a) Interoperability between neural network libraries like Caffe, TensorFlow; (b) Neural Network exchange formats; (c) Sensor resolution versus CNN resolution; (d) Reusability of AI code between different family of SoCs; (e) Traceability and verification of deep networks, and their impact on the SoC.

Bio: Vikram Narayan is currently heading the AI & Computer Vision Team at the ADAS Group of Visteon Corporation and based out of its corporate head quarter at Detroit. He is responsible for camera and LiDAR based sensing technologies and has delivered functions like lane keeping which has been tested on the DriveCore ECU, which is Visteon's centralized domain controller. In addition, the AI & Computer Vision Team of Visteon has filed several patent applications and published at premier venues like ECCV and IVS. He holds a Ph.D in Robotics from Uni Bielefeld, Germany and pursued his postdoctoral research at Frankfurt Institute of Advanced Studies, Germany. Prior to his current stint at Visteon, he worked at Panasonic Automotive EU for camera based automatic parking.

Visteon helped the industry move beyond hard buttons and knobs to touch screens and haptic controllers. Today, we're the first in the industry to address the proliferation of electronic control units (ECUs) through our SmartCore™ cockpit domain controller solution. Now, another transformation is underway, driven by requirements for autonomous vehicles. This new phase will be characterized by technology such as artificial intelligence, augmented reality, driver monitoring and interior sensing. We're participating in this new era of technology in the same place we've always been – the fast lane.

Talk 5: "Introduction to Face Processing with Computer Vision"

Mr. Gabriel Bianconi, Founder, Scalar Research

Abstract: Ever wonder how Facebook's facial recognition or Snapchat's filters work? In this talk, we'll help you understand some of the computer vision and machine learning techniques behind these applications and use this knowledge to develop our own prototypes for the products above. Faces are a fundamental piece of photography and building applications around them has never been easier with open-source libraries and pre-trained models. In this talk, we'll help you understand some of the computer vision and machine learning techniques behind these applications. Then, we'll use this knowledge to develop our own prototypes to tackle tasks such as face detection (e.g. digital cameras), recognition (e.g. Facebook Photos), classification (e.g. identifying emotions), manipulation (e.g. Snapchat filters), and more.

Bio: Gabriel Bianconi is a machine learning scientist with experience in applying cutting-edge academic research to solve real-world problems. He began his training as a B.S. & M.S. student in computer science at Stanford University, where he received multiple academic distinctions, including the President's Award for Academic Excellence. He was one of ten students to graduate with honors in computer science in his undergraduate class at Stanford. His thesis investigated quantum deep learning algorithms using NASA's D-Wave quantum computer and was selected for a presentation at the AQC 2017 Conference in Tokyo, Japan. During his master's program, he conducted research at the Stanford Partnership in AI-Assisted Care, a joint lab between the Stanford Computer Science Department (Prof. Fei-Fei Li – Chief Scientist of Cloud AI/ML at Google) and the Stanford School of Medicine (Prof. Arnold Milstein). His research focused on improving clinical care and reducing monitoring costs in hospitals by leveraging machine learning and computer vision and resulted in a first-author manuscript selected as Top 10 Research Paper at the NIPS Machine Learning for Health 2017 Workshop.

Gabriel also has extensive software engineering experience. At Google and Facebook, he worked on backend infrastructure for enterprise tools responsible for billions of dollars in revenue. He's also created an advertising supply-side platform that handled millions of ad requests per day, built an algorithmic trading platform and quantitative strategies for crypto asset markets that handled over US$10M in volume, and held positions at startups and investment firms.

Gabriel is the founder of Scalar Research, an artificial intelligence and data science consulting firm. Scalar Research help companies tackle complex business challenges with data-driven products and solutions leveraging cutting-edge machine learning and advanced analytics.

ValleyML.ai State of AI and ML-Spring 2019

Visit ValleyML.ai for other related events and hands-on workshops.

Dr. Kiran Gunnam, Program Chair, ValleyML.ai

Distinguished Engineer - Machine Learning & Computer Vision

Dr. Gunnam is an innovative technology leader with vision and passion who effectively connects with individuals and groups. Dr. Gunnam's breakthrough contributions are in the areas of advanced error correction systems, storage class memory systems and computer vision-based localization & navigation systems. He has helped drive organizations to become industry leaders through ground-breaking technologies. Dr. Gunnam has 75 issued patents and 100+ patent applications/invention disclosures on algorithms, architectures and real-time low-cost implementations for computing, storage and computer vision systems. He is the lead inventor/sole inventor for 90% of them. Dr. Gunnam's patented work has been already incorporated in more than 2 billion data storage and WiFi chips and is set to continue to be incorporated in more than 500 million chips per year.

Dr. Gunnam is also a key contributor to the precise localization and navigation technology commercialized for autonomous aerial refueling and space docking applications. His recent patent pending inventions on low-complexity simultaneous localization and mapping (SLAM) and 3D convolutional neural network (CNN) for object detection, tracking and classification are being commercialized for LiDAR+camera-based perception for autonomous driving and robotic systems.

Dr. Gunnam received his MSEE and PhD in Computer Engineering from Texas A&M University, College Station. He is world-renowned for balance between strong analytical ability and pragmatic insight into implementation of advanced technology. He served as IEEE Distinguished Speaker and Plenary Speaker for 25+ events and international conferences and more than 3000 attendees in USA, Canada and Asia benefited from his lecture talks. He also teaches graduate level course focused on machine learning systems at Santa Clara University.

ValleyML.ai State of AI and ML-Spring 2019

Visit ValleyML.ai for other related events and hands-on workshops.