

tinyML[®] Summit 2022

Enabling Ultra-low Power Machine Learning at the Edge

March 28-30, 2022 | Burlingame, California

Monday, March 28, 2022

tinyML Research Symposium

9:00 am – 9:30 am

Welcome and Opening Statement

9:30 am – 10:50 am

tinyML Applications and Systems

Session Co-Chairs: Zain Asgar, Co-Founder/CEO, Pixie and Wolfgang Furtner

Millimeter-Scale Ultra-Low-Power Imaging System for Intelligent Edge Monitoring

How to Manage Tiny Machine Learning at Scale - An Industrial Perspective

Distributed On-Sensor Compute System for AR/VR Devices: A Semi-Analytical Simulation Framework for Power Estimation

IMU Preintegrated Features for Efficient Deep Inertial Odometry

10:50 am – 11:10 am

Break

Session 1 - Short papers

11:10 am – 11:50 am

Session Chair: Amey Kulkarni, Senior Software Engineer, NVIDIA

A Semi-Decoupled Approach to Fast and Optimal Hardware-Software Co-Design of Neural

Combinatorial-Randomness-Based Power Amplifier Datasets with RF Fingerprint Classification

Combinatorial RL-based Scheduling for Pipelined Edge TPUs

A Fast Network Exploration Strategy to Profile Low Energy Consumption for Keyword Spotting

11:50 am – 12:50 pm

Lunch

Monday, March 28, 2022 (continued)

12:50 pm – 2:30 pm

tinyML Algorithms

Session Co-Chairs: Yingyan Lin, Assistant Professor, Rice University and Laura Galindez, Postdoctoral Researcher, UC Berkeley

LDP: Learnable Dynamic Precision for Efficient Deep Neural Network Training and Inference

PocketNN: Integer-only Training and Inference of Neural Networks via Direct Feedback Alignment and Pocket Activations in Pure C++

tinyMAN: Lightweight Energy Manager using Reinforcement Learning for Energy Harvesting Wearable IoT Devices

Delta Keyword Transformer: Bringing Transformers to the Edge through Dynamically Pruned Multi-Head Self-Attention

Toward Compact Deep Neural Networks via Energy-Aware Pruning

Break

2:30 pm – 2:50 pm

tinyML Hardware

3:50 pm – 4:30 pm

Improving the Energy Efficiency and Robustness of tinyML Computer Vision using Log-Gradient Input Images

An Empirical Study of Low Precision Quantization for TinyML

Power-of-Two Quantization for Low Bitwidth and Hardware Compliant Neural Networks

A Brain-Inspired Low-Dimensional Computing Classifier for Inference on Tiny Devices

TinyM²Net: A Flexible System Algorithm Co-designed Multimodal Learning Framework for Tiny Devices

Break

4:30 pm – 4:50 pm

Session 2 - Short papers

4:50 pm – 5:30 pm

L3U-Net: Low-Latency Lightweight U-Net Based Image Segmentation Model for Parallel CNN Processors

Neural Architecture Search for Energy Efficient Always-on Audio Models

Towards Agile Design of Neural Processing Units with Chisel

Neural Architecture Search for Low-Precision Neural Networks

Tuesday, March 29, 2022

Day 1 - tinyML Summit

8:00 am – 9:00 am

Registration and Breakfast

9:00 am – 9:15 am

Welcome/Introduction

9:15 am – 10:00 am

Opening Talk

Kate Kallot, Head of Global Development Relations & Emerging Areas, NVIDIA

10:00 am – 10:15 am

Break

10:15 am – 11:45 am

Session 1: tinyML Vision

Session Leader: Adam Fuks, NXP

10:15 - 10:45am

TinyML for All: Full-stack Optimization for Every Edge AI Platform

Di Wu, Co-founder and CEO OmniML

10:45 – 11:05 am

Tiny models with big appetites: Cultivating the perfect data diet

Jelmer Neeven, Deep learning scientist and software engineer, Plumerai

11:05 – 11:25 am

Enabling tiny camera sensors for Augmented Reality

Rakesh Renjen, Senior Research Scientist Manager in Meta Reality Labs, Facebook

11:25 – 11:45 am

NDP200 tinyML Vision Processing

Dave Garrett, Chief Architect, Syntiant

11:45 am – 1:15 pm

Interaction / Lunch

1:15 pm – 3:00 pm

Session 2: tinyML Audio

Session Leader: Chris Rowen, Cisco

1:15 - 1:20 pm

Intro Chris Rowen

1:20 – 1:45pm

On device speech models optimization and deployment

Oleg Rybakov, Software Engineering Manager, Google

1:45 – 2:10 pm

AnalogML: Analog Inferencing for System-Level Power Efficiency

David Graham, Co-founder and Chief Science Officer, Aspinity Inc.

2:10 – 2:35 pm

Dissecting a low power AI/ML edge application: Noise Suppression

Raj Pawate, Group Director, Tensilica IPG, Cadence Design Systems Inc.

2:35 – 3:00 pm

Real-time deep speech enhancement system for embedded voice UI

Tess Boivin, ML Software Engineer, NXP

3:00pm – 3:45pm Poster Session / Break

3:45 – 5:15 pm

Session 3: Sensing for tinyML

Session Moderator: Steve Whalley, CEO, Strategic World Ventures

3:45 – 4:10 pm

Brains into sensors with AI in the Edge

Andrea ONETTI, Executive Vice President, STMicroelectronics

4:10-4:35 pm

Sensing Applications as a Driver for TinyML Solutions

Victor PANKRATIUS, Director / Head of Global Software Engineering, Bosch Sensortec GmbH

4:35 – 5:00 pm

Sensors and ML: waking smarter for less

Peter HARTWELL, CTO, TDK InvenSense

5:00- 5:15 pm

Q&A

5:15pm – 7:15pm

Reception

March 30, 2022

Day 2 – tinyML Summit

8:00am - 9:00pm

Registration and Breakfast

9:00am - 9:15am

Welcome

9:15am – 11:10am

Hardware

Session Leader: Francesco Conti Assistant Professor, University of Bologna, Italy

9:15 – 9:40am

Perspectives & Challenges for TinyML Hardware: a System-Level View

Francesco Conti, Assistant Professor, University of Bologna, Italy

9:40 – 10:05am

Programmable In-Memory Computing (IMC) Accelerator with >100 SRAM IMC Macros

Jae-Sun Seo, Assistant Professor, ASU

10:05 – 10:20am

Break

10:20-10:45am

Mastering the 3 Pillars of AI Acceleration: Algorithms, Hardware and Software

Swagath VENKATARAMANI, Research Staff Member, IBM

10:45 – 11:10 am

Next-Generation Deep-Learning Accelerators: From Hardware to System

Sophia SHAO, Assistant Professor of Electrical Engineering and Computer Sciences, University of California, Berkeley

11:10am -11:55am

Posters / Exhibits

11:55am - 12:40am

Lunch

12:40pm - 1:25pm

Session 5: Software Session - Part 1

Session Leader: Alessandro Grande, Director of Technology, Edge Impulse

12:40 – 12:55pm

Ecosystem of tools for better productivity

Danilo PAU, Technical Director, IEEE and ST Fellow, STMicroelectronics Italia

12:55 – 1:10pm

Compiling TinyML Models with microTVM

Andrew REUSCH, Software Engineer, OctoML

1:10 – 1:25pm

Building data-centric AI tooling for embedded engineers

Daniel SITUNAYAKE, Founding tinyML Engineer, Edge Impulse

1:25-1:55pm

Break

1:55-2:40pm

Session 5: Software / Tools

Session Leader: Alessandro Grande, Director of Technology, Edge Impulse

1:55 – 2:10pm

Expedited Model Deployment on the Edge with Recipes

Haya Sridharan, Technical Product Manager, Latent AI

2:10 – 2:25pm

Suitability of TinyML for addressing predictive maintenance in high tech manufacturing

Christopher Knorowski, CTO, SensiML Corp

2:25 – 2:40pm

Challenges for Large Scale Deployment of Tiny ML Devices

Gopal Raghavan, Embedded AI Strategy, Microsoft

2:40pm -3:10pm

Summit Wrap up / break

3:10pm - 5:30pm

Auto tinyML

Session Leader: Danilo PAU, Technical Director, IEEE and ST Fellow, STMicroelectronics Italia

3:10 - 3:35pm

EON Tuner: AutoML for constrained devices

Jan Jongboom, Edge Impulse

3:35 – 4:00pm

Optimizing AutoML for the tinyML Future

Elias FALLON, Vice President for Machine Learning, Qeexo Co.

4:00 – 4:15pm

Break

4:15 – 4:40pm

1 kB and not a bit more! The ideal weight for a tinyML model

Blair Newman, CTO, Neuton

4:40 – 5:05pm

Model Optimization with QKeras' Quantization-Aware Training and Vizier's Automatic Neural Architecture Search

Daniele Moro, Software Engineer, Google

5:05 – 5:30pm

Automated Machine Learning under model's deployability on tiny devices

Antonio CANDELIERI, Assistant Professor, University of Milano-Bicocca, Italy

5:30 – 5:55pm

Automating Model Optimization for Efficient Edge AI: from automated solutions to open-source toolkit

Dave CHENG, Senior Deep Learning Researcher, Qualcomm AI Research