tinyML EMEA Technical Forum 2021 Proceedings

June 7 – 10, 2021
Virtual Event

www.tinyML.org
TOD: Transprecise Object Detection to Maximise Real-Time Accuracy on the Edge

Presented by: JunKyu Lee, Research Fellow, Queen’s University Belfast, UK

Acknowledgement
The European Commission Horizon 2020 No. 732631 (OPRECOMP)
The Engineering and Physical Sciences Research Council No. EP/T022345/1 (DiPET)
The CHIST-ERA No. CHIST-ERA-18-SDCDN-002 (DiPET)

DiPET Website: https://dipet.eeecs.qub.ac.uk

June 10, 2021
- Edge for Real-Time Video Analytics (RTVA)

- Challenges in RTVA on the Edge: Limited Computational Power → Longer Inference Latency from Deep Neural Network (DNN) → More Dropped Frames → Limited Object Detection Accuracy

Figure Source: Multi Object Tracking 17 Dataset (https://motchallenge.net)
TOD: Transprecise Object Detection (to maximise RTVA accuracy on the Edge)

- Initialization: Preloading 4 different YOLOv4 detectors into GPU RAM
- Runtime Scheduler (RS) Selects a DNN on the Fly according to Dynamic Object Sizes Detected.
- TOD Demonstration when objects are fading away from camera (i.e., objects are getting smaller.)

Figure Source: Multi Object Tracking 17 Dataset (https://motchallenge.net)
Experimental Evaluation with a NVidia Jetson Nano Board with MAX power mode

Best (or equivalently good) accuracy for all cases
- 35% gain over YOLOv4-tiny-288.

With MOT17-05 data, TOD requires 63% of the power without losing accuracy compared to YOLOv4-416. (TOD: 5.7W, YOLOv4-416: 7.5W)
Automated TinyML

Zero-code SaaS solution

Create tiny models, ready for embedding, in just a few clicks!

Compare the benchmarks of our compact models to those of TensorFlow and other leading neural network frameworks.

Executive Sponsors
Arm: The Software and Hardware Foundation for tinyML

1. Connect to high-level frameworks
2. Supported by end-to-end tooling
3. Connect to Runtime

Profiling and debugging tooling such as Arm Keil MDK

Application

Optimized models for embedded

Runtime (e.g. TensorFlow Lite Micro)

Optimized low-level NN libraries (i.e. CMSIS-NN)

RTOS such as Mbed OS

Arm Cortex-M CPUs and microNPUs

Stay Connected

@ArmSoftwareDevelopers

@ArmSoftwareDev

Resources: developer.arm.com/solutions/machine-learning-on-arm
TinyML for all developers

C++ library
Arduino library
WebAssembly

Dataset
Acquire valuable training data securely
Enrich data and train ML algorithms

Edge Device
Real sensors in real time
Open source SDK

Embedded and edge compute deployment options

Test
Test impulse with real-time device data flows

www.edgeimpulse.com
Qualcomm AI Research is an initiative of Qualcomm Technologies, Inc.

Advancing AI research to make efficient AI ubiquitous

Power efficiency
- Model design, compression, quantization, algorithms, efficient hardware, software tool

Personalization
- Continuous learning, contextual, always-on, privacy-preserved, distributed learning

Efficient learning
- Robust learning through minimal data, unsupervised learning, on-device learning

Perception
- Object detection, speech recognition, contextual fusion

Reasoning
- Scene understanding, language understanding, behavior prediction

Action
- Reinforcement learning for decision making

A platform to scale AI across the industry
Syntiant Corp. is moving artificial intelligence and machine learning from the cloud to edge devices. Syntiant’s chip solutions merge deep learning with semiconductor design to produce ultra-low-power, high performance, deep neural network processors. These network processors enable always-on applications in battery-powered devices, such as smartphones, smart speakers, earbuds, hearing aids, and laptops. Syntiant's Neural Decision Processors™ offer wake word, command word, and event detection in a chip for always-on voice and sensor applications.

Founded in 2017 and headquartered in Irvine, California, the company is backed by Amazon, Applied Materials, Atlantic Bridge Capital, Bosch, Intel Capital, Microsoft, Motorola, and others. Syntiant was recently named a CES® 2021 Best of Innovation Awards Honoree, shipped over 10M units worldwide, and unveiled the NDP120 part of the NDP10x family of inference engines for low-power applications.

www.syntiant.com

@Syntiantcorp
Platinum Sponsors
Part of your life. Part of tomorrow.

www.infineon.com
Add Advanced Sensing to your Product with Edge AI / TinyML

Pre-built Edge AI sensing modules, plus tools to build your own

Reality AI solutions
- Prebuilt sound recognition models for indoor and outdoor use cases
- Solution for industrial anomaly detection
- Pre-built automotive solution that lets cars "see with sound"

Reality AI Tools® software
- Build prototypes, then turn them into real products
- Explain ML models and relate the function to the physics
- Optimize the hardware, including sensor selection and placement

https://reality.ai  info@reality.ai  @SensorAI  Reality AI
Gold Sponsors
Adaptive AI for the Intelligent Edge

LatentAI.com
Build Smart IoT Sensor Devices From Data

SensiML pioneered TinyML software tools that auto generate AI code for the intelligent edge.

- End-to-end AI workflow
- Multi-user auto-labeling of time-series data
- Code transparency and customization at each step in the pipeline

We enable the creation of production-grade smart sensor devices.

sensiml.com
Silver Sponsors

- emza
- GREENWAVES TECHNOLOGIES
- HOTG
- imagimob
- seeed
- life.augmented
Copyright Notice

The presentation(s) in this publication comprise the proceedings of tinyML® EMEA Technical Forum 2021. The content reflects the opinion of the authors and their respective companies. This version of the presentation may differ from the version that was presented at tinyML EMEA. The inclusion of presentations in this publication does not constitute an endorsement by tinyML Foundation or the sponsors.

There is no copyright protection claimed by this publication. However, each presentation is the work of the authors and their respective companies and may contain copyrighted material. As such, it is strongly encouraged that any use reflect proper acknowledgement to the appropriate source. Any questions regarding the use of any materials presented should be directed to the author(s) or their companies.

tinyML is a registered trademark of the tinyML Foundation.

www.tinyML.org