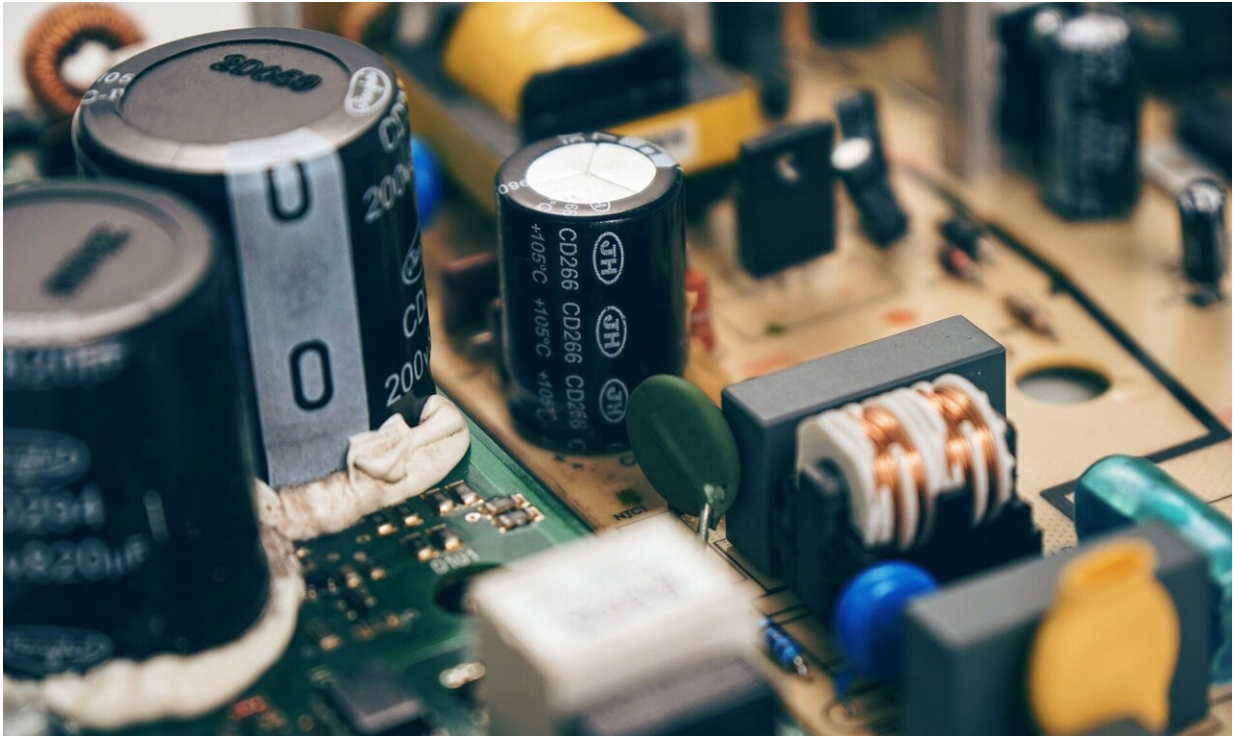


Setting the standard for machine learning

June 25 2019, by Leah Burrows



Credit: CC0 Public Domain

The microcomputer revolution of the 1970s triggered a Wild West-like expansion of personal computers in the 1980s. Over the course of the decade, dozens of personal computing devices, from Atari to Xerox Alto, flooded into the market. CPUs and microprocessors advanced rapidly, with new generations coming out on a monthly basis.

Amidst all that growth, there was no standard method to compare one computer's performance against another. Without this, not only would consumers not know which system was better for their needs but computer designers didn't have a standard method to test their systems.

That changed in 1988, when the Standard Performance Evaluation Corporation (SPEC) was established to produce, maintain and endorse a standardized set of performance benchmarks for computers. Think of benchmarks like standardized tests for computers. Like the SATs or TOEFL, benchmarks are meant to provide a method of comparison between similar participants by asking them to perform the same tasks.

Since SPEC, dozens of benchmarking organizations have sprung up to provide a method of comparing the performance of various systems across different chip and program architecture.

Today, there is a new Wild West in [machine learning](#). Currently, there are at least 40 different hardware companies poised to break ground in new AI processor architectures.

"Some of these companies will rise but many will fall," said Vijay Janapa Reddi, Associate Professor of Electrical Engineering at the Harvard John A. Paulson School of Engineering and Applied Sciences (SEAS). "The challenge is how can we tell if one piece of hardware is better than another? That's where benchmark standards become important."

Janapa Reddi is one of the leaders of [MLPerf](#), a machine learning benchmarking suite. ML Perf began as a collaboration between researchers at Baidu, Berkeley, Google, Harvard, and Stanford and has grown to include many companies, a host of universities, along with hundreds of individual participants worldwide. Other Harvard contributors include David Brooks, the Haley Family Professor of

Computer Science at SEAS and Gu-Yeon Wei, the Robert and Suzanne Case Professor of Electrical Engineering and Computer Science at SEAS.

The goal of ML Perf is to create a benchmark for measuring the performance of machine learning software frameworks, machine learning hardware accelerators, and machine learning cloud and edge computing platforms.

We spoke to Janapa Reddi about MLPerf and the future of benchmarking for machine learning.

SEAS: First, how does benchmarking for machine learning work?

Janapa Reddi: In its simplest form, a benchmark standard is a strict definition of a machine learning task, let's say image classification. Using a model that implements that task, such as ResNet50, and a dataset, such as COCO or ImageNet, the model is evaluated with a target accuracy or quality metric that it must achieve when it is executed with the dataset.

SEAS: How does benchmarking factor into your research at SEAS?

Janapa Reddi: Personally, I am interested in benchmarking autonomous and "tiny" machine learning systems.

Autonomous vehicles rely heavily on machine learning for vision processing, sensor fusion and more. The trunk of an autonomous car contains over 2,500 Watts of compute horsepower. Just to put that into context, a smartphone uses 3 Watts, and your average laptop uses 25

Watts. So these autonomous vehicles consume a significant amount of power, thanks in part to all the machine learning they rely upon. My Edge Computing Lab is interested in cutting down that power consumption, while still pushing the limits of all the processing capabilities that is needed, machine learning and all included.

At the other end of the spectrum are "tiny" devices. Think tiny little microcontrollers that consume milliwatts in power that can be tossed around and forgotten. Tiny microcontrollers today are passive devices with little to no on-board intelligence. But "TinyML" is an emerging concept that focuses on machine learning for tiny embedded microcontrollers. My group is studying how we can enable TinyML since we see many diverse uses. TinyML devices can monitor your health intelligently, or tiny drones that fit in your palm can navigate through tight small spaces in the event of a fallen building for search and rescue operations, and fly in between trees and leaves to monitor the health of farmer's crops and keep pests out

These are two domains that greatly interest me, specifically in the context of machine learning systems, because there are several interesting research problems to solve that extend beyond just machine learning hardware performance and include machine learning system software design and implementation.

SEAS: What lessons can machine learning take from previous benchmarking efforts, such as those started by SPEC three decades ago?

Janapa Reddi: Over the years, SPEC CPU has been driven by a consortium of different industry partners who come together to determine a suite of workloads that can lead to fair and useful benchmarking results. Hence, SPEC workloads have become a standard

in research and academia for measuring and comparing CPU performance. As David Patterson—a renowned computer architect and the 2017 Turing Award recipient—often likes to point out, SPEC workloads led to the golden age of microprocessor design.

We can borrow some lessons from SPEC and apply them toward machine learning. We need to bring the academic and research community together to create a similar consortium of industry partners who can help define standards and benchmarks that are representative of real-world use cases.

SEAS: Is that how ML Perf works?

Janapa Reddi: Yes. MLPerf is the effort of many organizations and several committed individuals, all working together with the single coherent vision of building a fair and useful benchmark for machine learning systems. Because of this team effort, we come up with benchmarks that are based on the wisdom of many people and a deep understanding of customer use cases from the real world. Engineers working on machine learning systems contribute their experiences with the nuanced systems issues and corporations can provide their real-world use cases (with user permission, of course). On the basis of all the information we gather, the MLPerf collaborative team of researchers and engineers curates a [benchmark](#) that is useful for learning platforms and systems.

SEAS: MLPerf just announced some new benchmarks for machine learning, right?

Janapa Reddi: Right. We've just announced [our first inference suite](#), which consists of five benchmarks across three different machine learning tasks: image classification, object detection and machine

translation. These three tasks include well known models like MobileNets and ResNet that support different image resolutions for different use cases like autonomous vehicles and smartphones.

We stimulate the models with the "LoadGen," which is a load generator that mimics different use case modes found in the real world. For instance, in smartphones, we take a picture, feed it into a machine learning model, and eagerly wait to see if it can identify what the image is. Obviously, we want that inference to be as fast as possible. In a camera monitoring system, we want to look at multiple pictures coming through different cameras, so the use case is sensitive to both latency and throughput (how many pictures can I process within a bounded amount of time). This LoadGen with our benchmarks sets MLPerf apart from other benchmarks.

SEAS: So, what comes next?

Janapa Reddi: Benchmarks are a step toward a bigger goal. MLPerf is interested in expanding its effort from curating benchmarks for evaluating system performance to developing new datasets that can foster new innovation in the machine learning algorithms, software and hardware communities. Thus far, we have been relying on datasets that have been largely made accessible via academics in the open source communities. But in some domains, like speech, there is a real need to develop new datasets that are at least 10 to 100 times larger. But bigger alone is insufficient. We also need to address fairness and the lack of diversity in the datasets to ensure that the models that are trained on these datasets are unbiased

SEAS: How are you addressing fairness and diversity in machine learning?

Janapa Reddi: We created "Harvard MLPerf Research" in conjunction with the Center for Research on Computation and Society (CRCS), which brings together scientists and scholars from a range of fields to make advances in computational research that serve public interest. Through the center, we hope to connect with the experts in other schools to address issues such as fairness and bias in datasets. We need more than computer scientists to address these issues.

Provided by Harvard University

Citation: Setting the standard for machine learning (2019, June 25) retrieved 19 April 2024 from <https://phys.org/news/2019-06-standard-machine.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.