RobotPerf: Benchmarking Robotic Systems

Harvard University

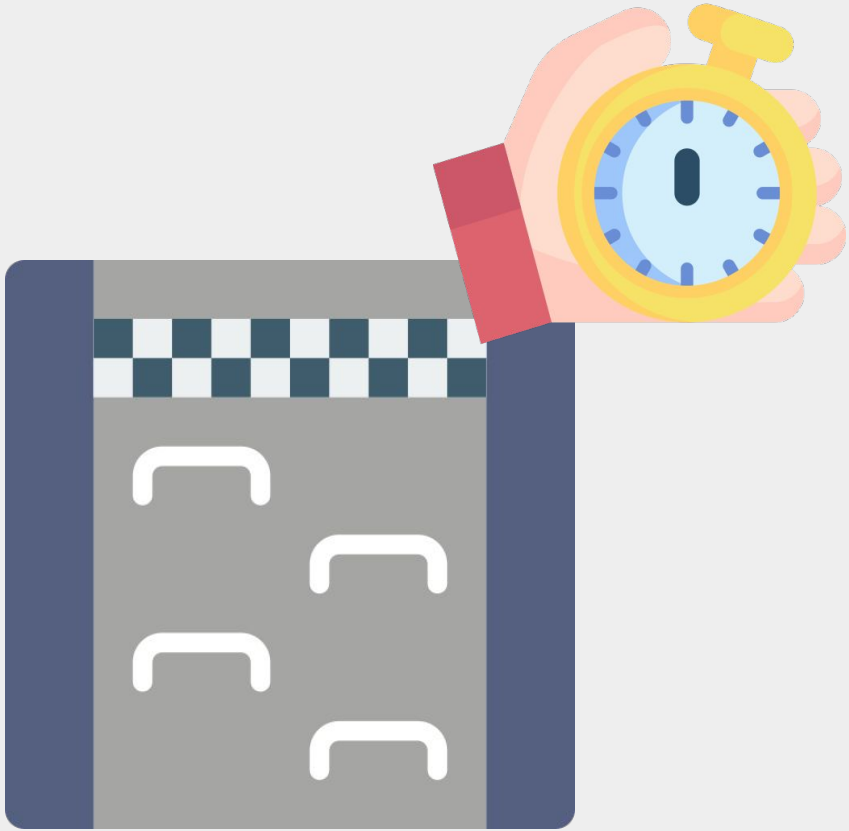
ACCELERATION ROBOTICS

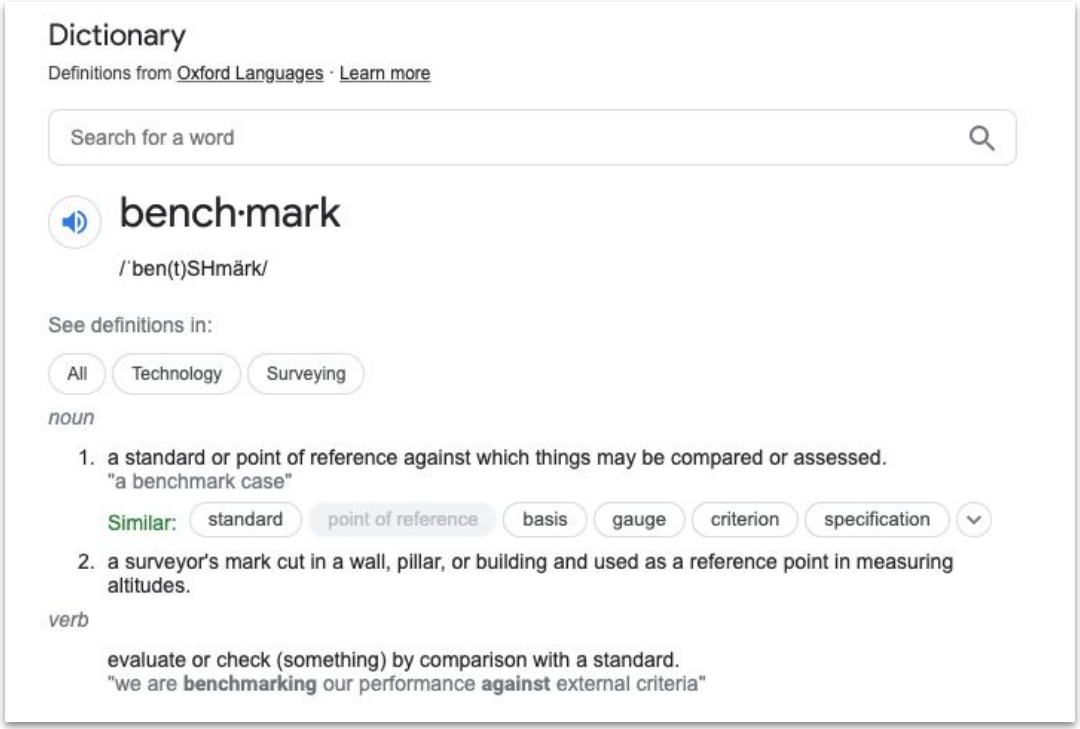
Prof. Vijay Janapa Reddi

Jason Jabbour

Víctor Mayoral-Vilches

What is benchmarking and why is it important?

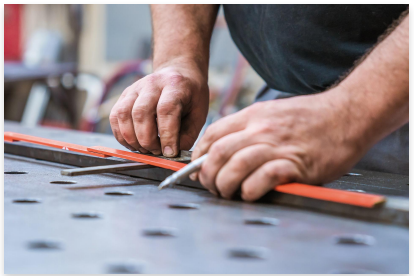



Dictionary
Definitions from Oxford Languages · Learn more
Search for a word
bench·mark
/ˈben(t)SHmärk/
See definitions in:
All
Technology
Surveying
noun
1. a standard or point of reference against which things may be compared or assessed.
"a benchmark case"
Similar:
standard
point of reference
basis
gauge
criterion
specification
2. a surveyor's mark cut in a wall, pillar, or building and used as a reference point in measuring altitudes.
verb
evaluate or check (something) by comparison with a standard.
"we are benchmarking our performance against external criteria"

Benchmarks

Use to

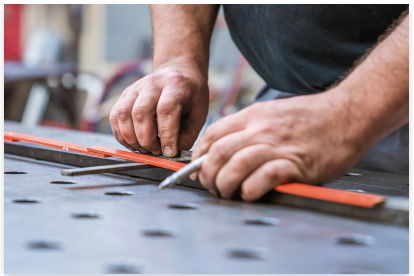
- **Compare** solutions
- **Inform** selection
- **Measure** and track progress
- **Raise** the bar**, advance** the field



Benchmarks

Use to

- **Compare** solutions
- **Inform** selection
- **Measure** and track progress
- **Raise** the bar**, advance** the field



Requires

- **Methodology** that is both fair and rigorous
- **Community** support and consensus

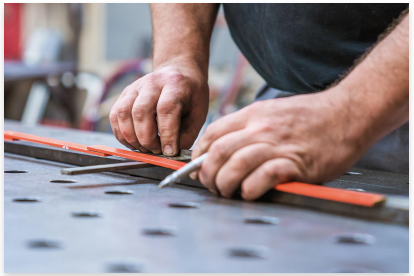
Benchmarks

Use to

- **Compare** solutions
- **Inform** selection
- **Measure** and track progress
- **Raise** the bar**, advance** the field

Requires

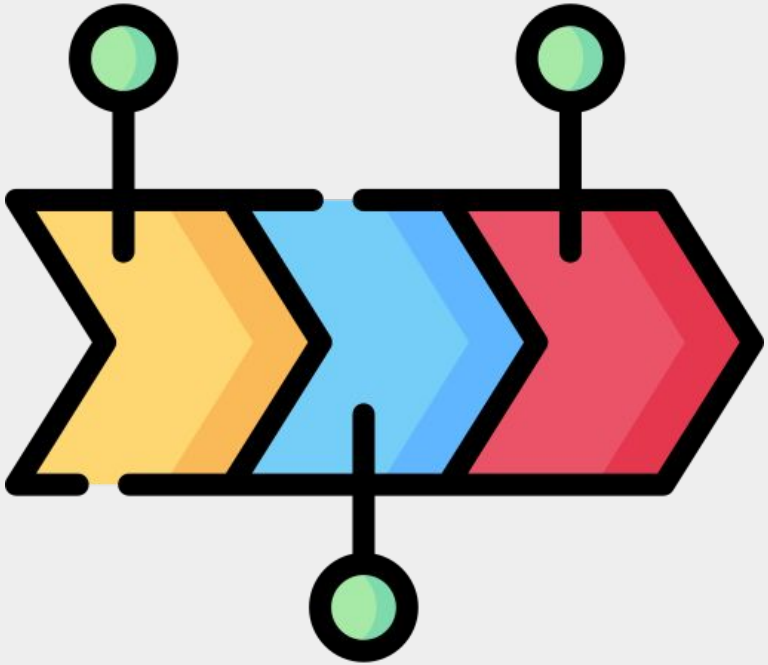
- **Methodology** that is both fair and rigorous
- **Community** support and consensus



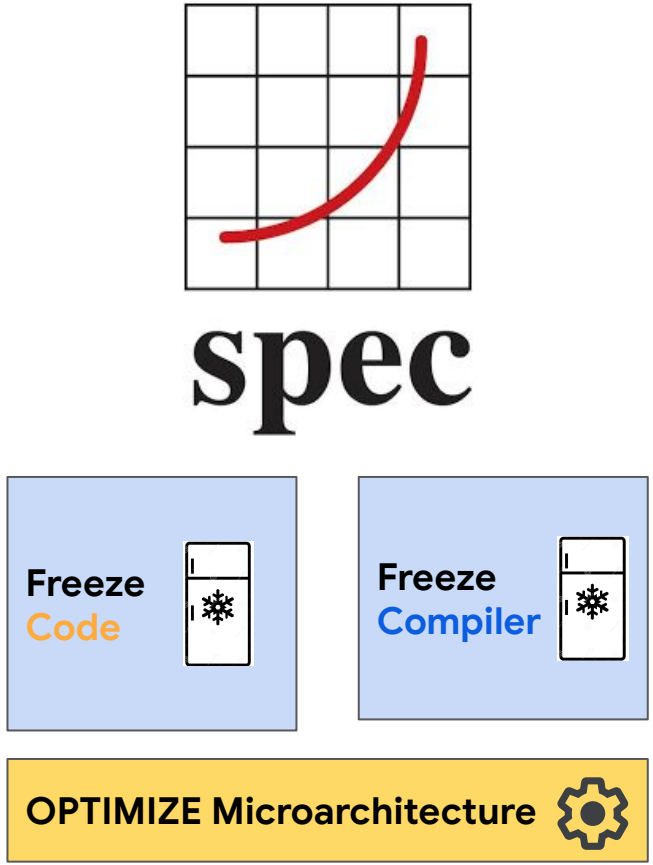
Provides

- **Standardization** of use cases and workloads
- **Comparability** across heterogeneous HW/SW systems
- **Complex characterization** of system compromises
- **Verifiable and Reproducible** results

Past, Present and Future of Benchmarking Systems



Widely-accepted benchmarks for **general-purpose CPUs**



Enforce performance result replicability to ensure reliable results

Use **representative workloads**, reflecting production use-cases

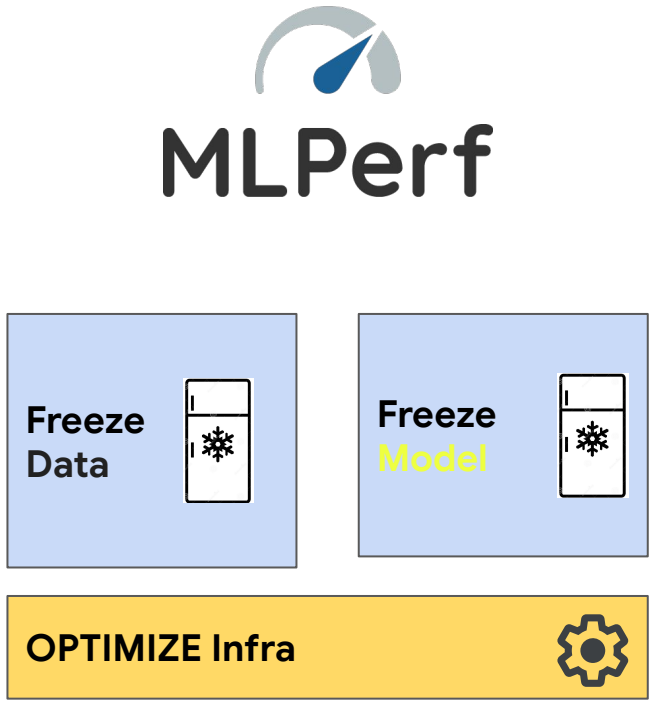
Encourage innovation to improve the state-of-the-art CPUs

Accelerate progress in CPUs via **fair and useful measurement**

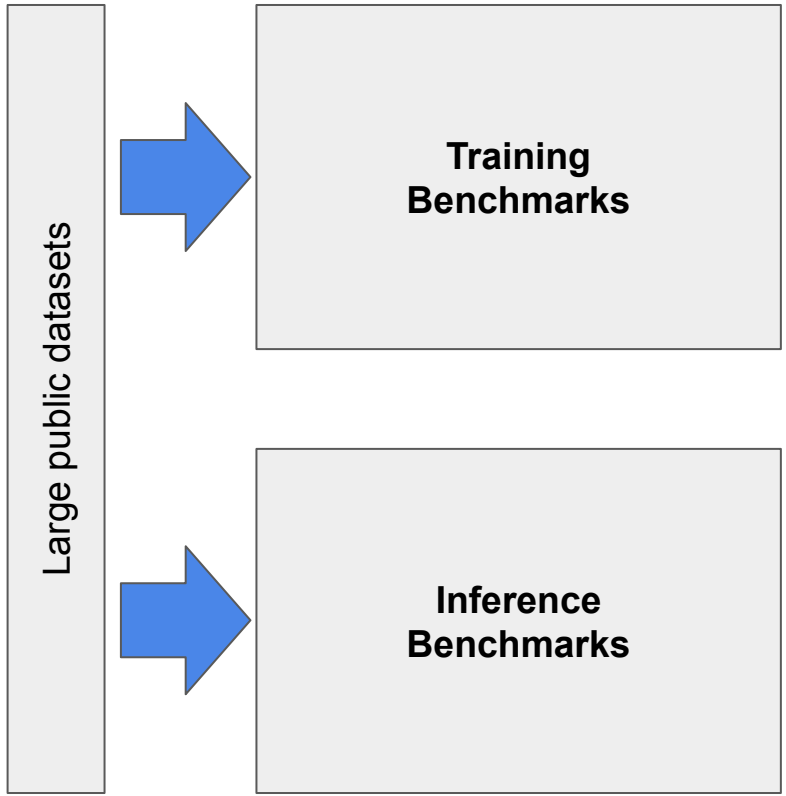
Serve the **commercial and research communities**

Keep **benchmarking affordable** so that all can participate

Widely-accepted benchmarks for **ML systems**



MLPerf



1. Identify a set of **ML tasks** and models
2. Identify real-world **scenarios** to emulate
3. Outline the **rules** for benchmarking
4. Define a clear set of evaluation **metrics**
5. Collect **results** to publish

A Community-driven **ML Benchmark Suite**

Harvard University, Stanford University, University of Arkansas, Littlerock, University of California, Berkeley, University of Illinois, Urbana Champaign, University of Minnesota, University of Texas, Austin

University of Toronto

AI Labs.tw, Alibaba, AMD, Andes Technology, Arm, Baidu, Cadence, Intel, In-Q-Tel, Lenovo, MediaTek, Mentor Graphics, Microsoft, Myrtle

Calypso AI, Centaur Technology, Cerebras, Ceva, Cisco, Cray, CTuning Foundation, Mythic, NetApp, NVIDIA, One Convergence, PathPartner Technology, Qualcomm, Rpa2ai

Dividiti, DDN Storage, Edgify, Enflame Tech, Esperanto, Facebook, Google, Sambanova, Samsung S.LSI, Sigopt, Synopsys, Tencent, Tensyr, Teradyne

Groq, Habana, Hop Labs, Horizon Robotics, Iluvatar, Inspur, Inzone AI, Transpire Ventures, VMware, Volley, Wave Computing, Wiwynn, WekaIO, Xilinx

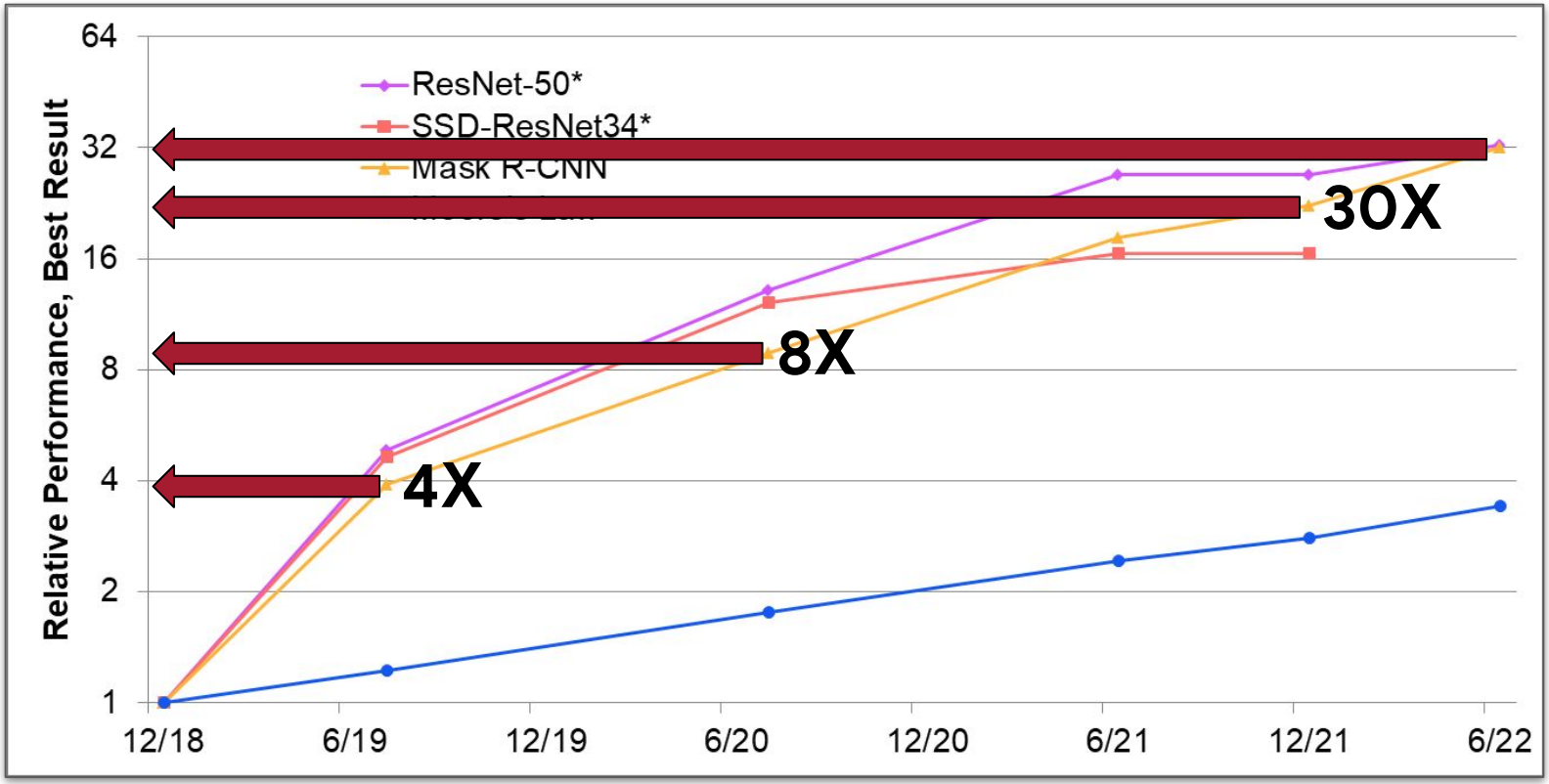
1,000+ members, 50+ organizations, 8+ universities

MLPerf Training Benchmark Snapshot (0.7v)

Area	Benchmark	Dataset	Quality Target	Reference Implementation Model
Vision	Image classification	ImageNet	75.90% classification	ResNet-50 v1.5
Vision	Object detection (light weight)	COCO	23.0% mAP	SSD
Vision	Object detection (heavy weight)	COCO	0.377 Box min AP and 0.339 Mask min AP	Mask R-CNN
Language	Translation (recurrent)	WMT English-German	24.0 Sacre BLEU	NMT
Language	Translation (non-recurrent)	WMT English-German	25.00 BLEU	Transformer
Language	NLP	Wikipedia 2020/01/01	0.712 Mask-LM accuracy	BERT
Commerce	Recommendation	1TB Click Logs	0.8025 AUC	DLRM

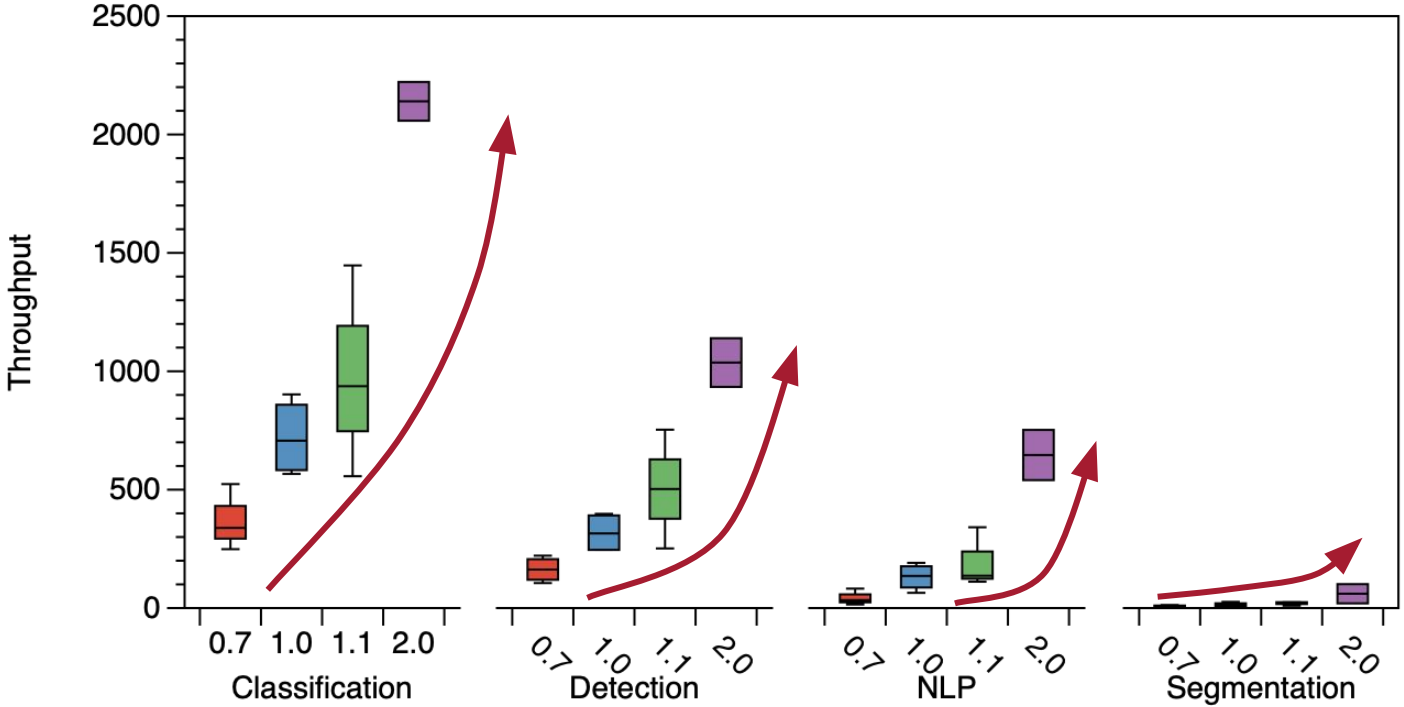
Cloud, Datacenter and Edge Inference

v0.5	v0.7	v1.0	Area	Task	Reference Model	Data Set	Quality Target (Top-1 Accuracy)
✔	✔	✔	**Vision**	Image Classification (Heavy)	ResNet-50 v1.5	ImageNet((224x224)	99% of FP32
✔	✔	✔	**Vision**	Object Detection (Heavy)	SSD-ResNet34	COCO (1,200 x 1,200)	99% of FP32
✔	✔	✔	**Vision**	Object Detection (Light)	SSD-MoblieNet-v1	COCO (300x300)	99% of FP32
✔			**Language**	Machine Translation	GNMT	WMT16 EN-DE	99% of FP32
	✔	✔	**Commerce**	Recommendation	DLRM	1TB Click Logs	99% of FP32 and 99.9% of FP32
	✔	✔	**Language**	Language Processing	BERT	SQuAD v1.1 (MAX_SEQ_LEN=384)	99% of FP32 and 99.9% of FP32
	✔	✔	**Speech**	Speech-to-Text	RNN-T	LibriSpeech Dev-Clean (Samples 15 seconds)	99% of FP32
	✔	✔	**Vision**	Medical Image Segmentation	3D U-Net	BRaTS 2019 (224x224x160)	99% of FP32 and 99.9% of FP32

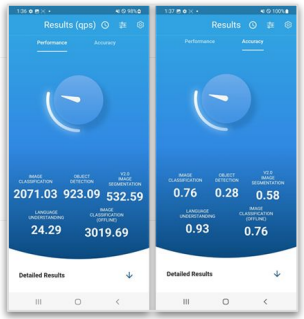


*indicates benchmark increased accuracy target in MLPerf **Training** v0.6

MLPerf Mobile **Inference** Performance Improvement



MLPerf Mobile



ML Commons


Alibaba Group
阿里巴巴集团
AMD
arm
Baidu 百度
CenTaur technology
cerebras
cisco
DELL EMC
Enflame
FACEBOOK AI
FURIOSA
GIGABYTE
Google
GML
GrAI Matter Labs
GRAPHCORE
地平线
Horizon Robotics
Hewlett Packard Enterprise
inspur
intel AI
KALRAY
LANDING AI
MEDIATEK
Microsoft
myrtle.ai
Nettrix 宁畅
NVIDIA
oppo
Qualcomm
Red Hat
SambaNova SYSTEMS
SAMSUNG Exynos
SYNTIANT
Tenstorrent
vmware
VMind
XILINX

MLCommons® is a Global Community

Members

Academics from educational institutions including:

Harvard University
Polytechnique Montreal
Peng Cheng Laboratory
Stanford University
University of California, Berkeley
University of Toronto
University of Tübingen
University of Virginia
University of York, United Kingdom
Yonsei University
York University, Canada

MLCommons aims to accelerate machine learning innovation to benefit everyone.

MLCommons aims to accelerate machine learning innovation to benefit everyone. Machine learning has tremendous potential to save lives in areas like healthcare and automotive safety and to improve information access and understanding through technologies like voice interfaces, automatic translation, and natural language processing. However, machine learning is completely unlike conventional software -- developers train an application rather than program it -- and requires a whole new set of techniques analogous to the breakthroughs in precision measurement, raw materials, and manufacturing that drove the industrial revolution.

MLCommons aims to answer the needs of the nascent machine learning industry through open, collaborative engineering in three areas:

Benchmarking

Benchmarks provide consistent measurements of accuracy, speed, and efficiency. Consistent measurements enable engineers to design reliable products and services, and enable researchers to compare innovations and choose the best ideas to drive the solutions of tomorrow.

Datasets

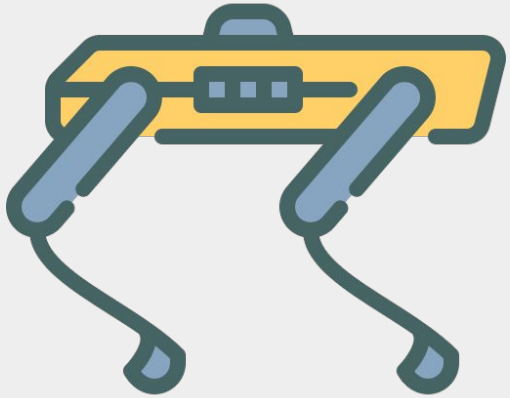
Datasets are the raw materials for all of machine learning. Models are only as good as the data they are trained on. Academics and entrepreneurs in particular depend on public datasets to create new technologies and new companies.

Best Practices

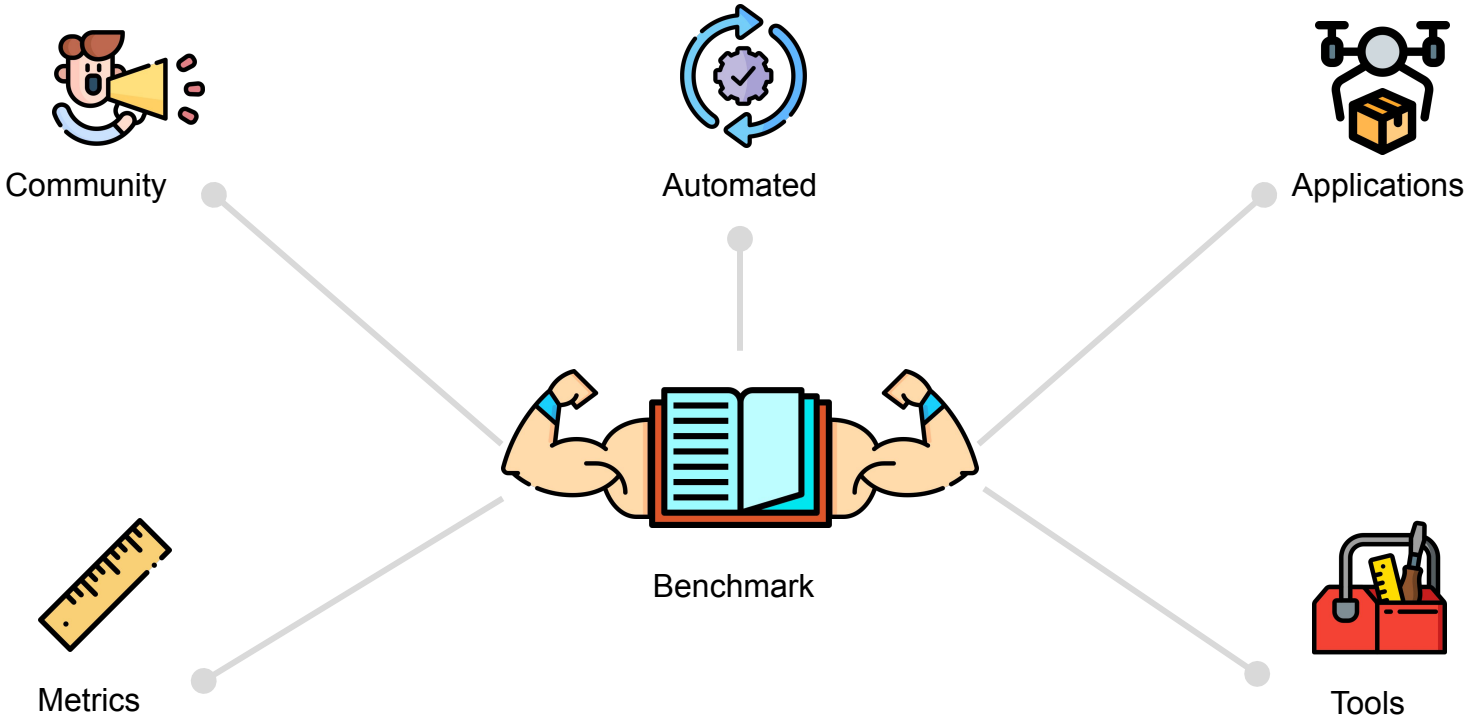
Best Practices empower researchers and engineers to more easily exchange models, reproduce experiments, and build applications that leverages machine learning. Improving best practices accelerates progress in, and grows the market for, machine learning.

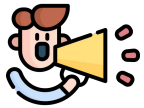
Can we reproduce the success of MLPerf and SPEC benchmarks for Robotics?

Benchmarking Robotic Systems



Characteristics of a Strong Benchmark





Listening to the Robotics Community

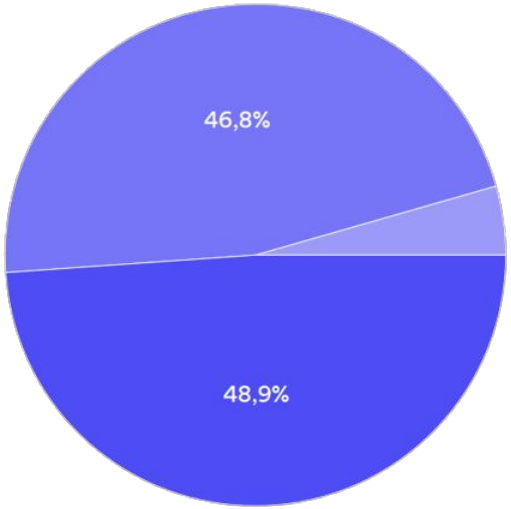
Acceleration Robotics. (2022). Hardware Acceleration Report in Robotics. https://accelerationrobotics.com/pdf/hardware_acceleration_robotics_2022.pdf

Valuable Metrics in Robotics

What do you care more about when it comes to hardware acceleration?

Valuable Metrics in Robotics

What do you care more about when it comes to hardware acceleration?

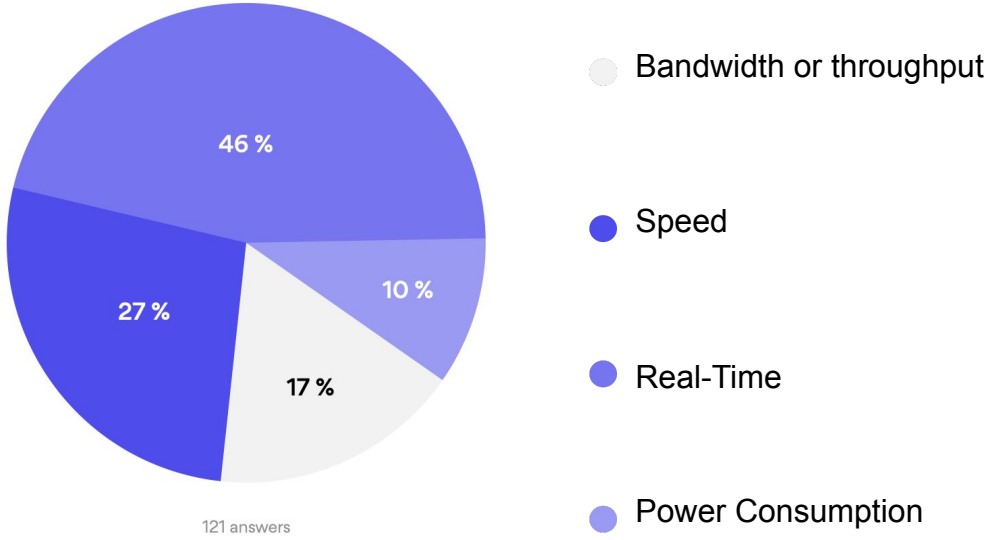


- Speed
- Real-Time & Determinism
- Power Consumption

Acceleration Robotics. (2022). Hardware Acceleration Report in Robotics. https://accelerationrobotics.com/pdf/hardware_acceleration_robotics_2022.pdf - 94 answers

Valuable Metrics in Robotics

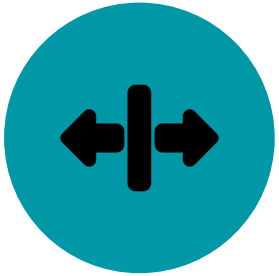
What do you care more about when it comes to hardware acceleration?



ROSCon 2022. Poll during talk. *What do you care more about when it comes to hardware acceleration? - 121 answers*

Valuable Metrics in Robotics

Latency



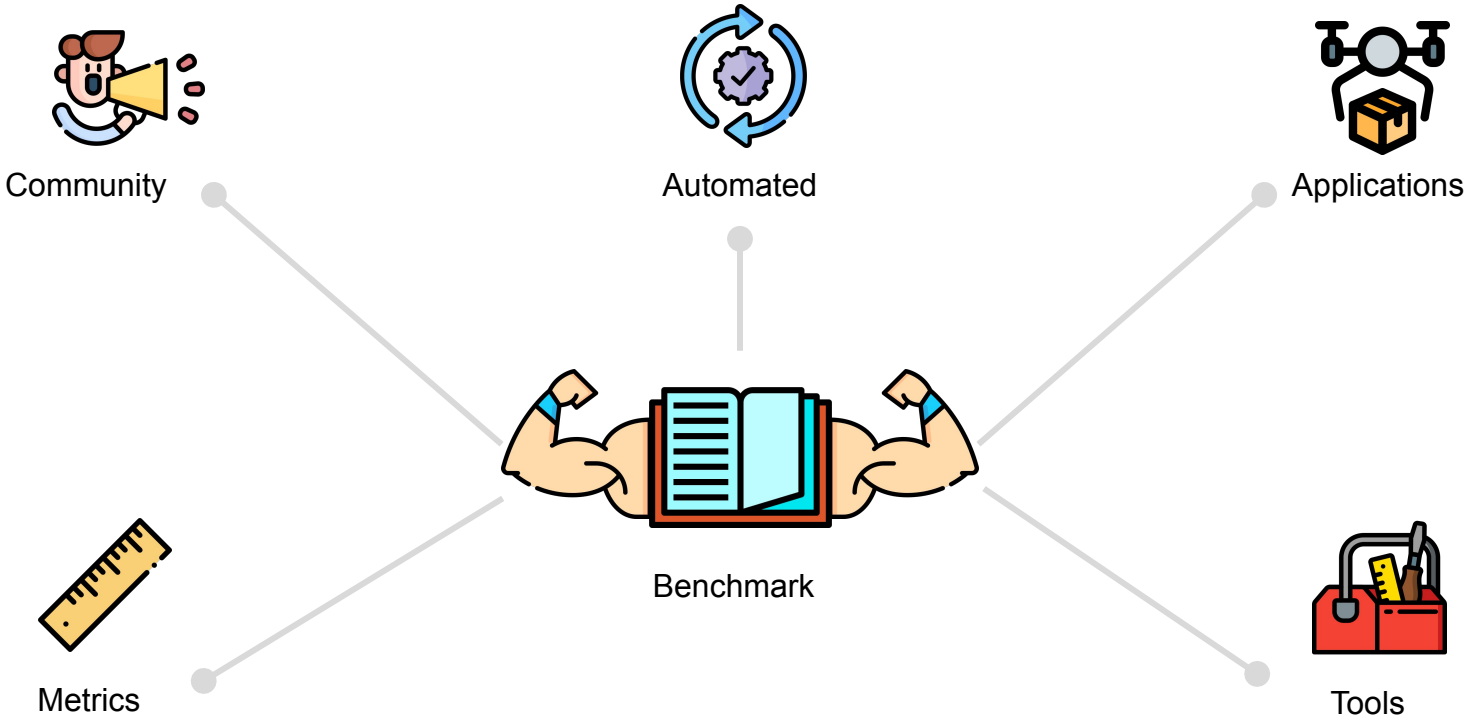
Throughput

Real-Time

Determinism

Power

Characteristics of a Strong Benchmark



Commonly used Tools

README.md

ros robotics companies 436 | acquired ros robotics companies 13

ROS Robotics Companies

Active companies | Acquired, closed or inactive | Contribute

A public list of companies that are known to use the Robot Operating System (ROS and ROS 2) or any of its related tools for development, to create products or to offer services. Ordered alphabetically.

Active companies

Company	Description	Year Founded
AATB	Design studio developing interactive robotic artworks and creative installations in the cultural, retail and entertainment sectors, also active in the film industry under the moniker Superposition. Using ROS on most projects since 2021.	2017
	Hardware Acceleration solutions for robots using ROS 2. Robot-specific	

100+ Companies

Commonly used Tools

100+ Companies

Commonly used Tools

ROS

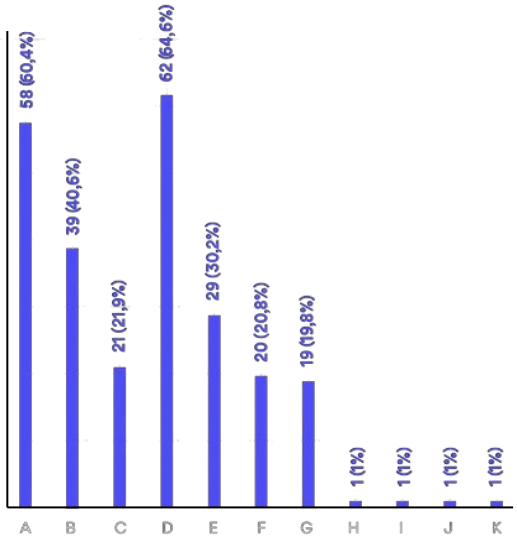
- Common Baseline
- Easy & Fast Development
- Prioritizes Performance

Commonly used Tools

What packages/components do you think we should prioritize when it comes to hardware accelerating ROS and/or Gazebo?

Commonly used Tools

What packages/components do you think we should prioritize when it comes to hardware accelerating ROS and/or Gazebo?

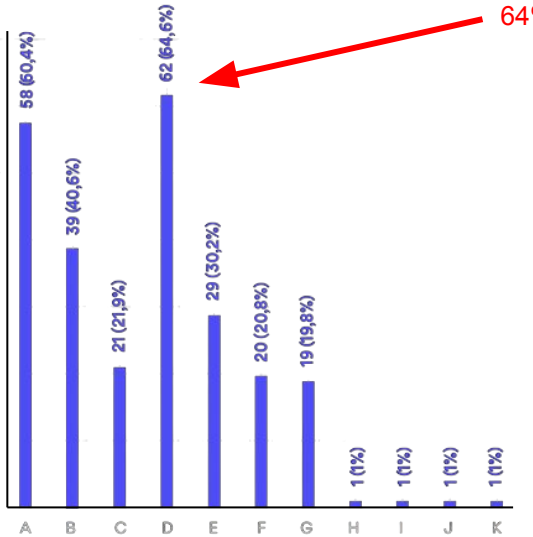


A. Gazebo/Ignition physic engines
B. ROS 2 navigation stack (navigation2)
C. ROS 2 manipulation stack (MoveIt2)
D. ROS 2 perception stack
E. ROS 2 communication middleware (DDS)
F. ROS 2 networking stack (UDP/IP/Ethernet, more deterministic network interactions)
G. ROS 2 control stack
H. All of them are important.
I. Webots physics engine
J. Image and depth data processing pipelines.
K. Lidar drivers and perception

Acceleration Robotics. (2022). Hardware Acceleration Report in Robotics. https://accelerationrobotics.com/pdf/hardware_acceleration_robotics_2022.pdf

Commonly used Tools

What packages/components do you think we should prioritize when it comes to hardware accelerating ROS and/or Gazebo?

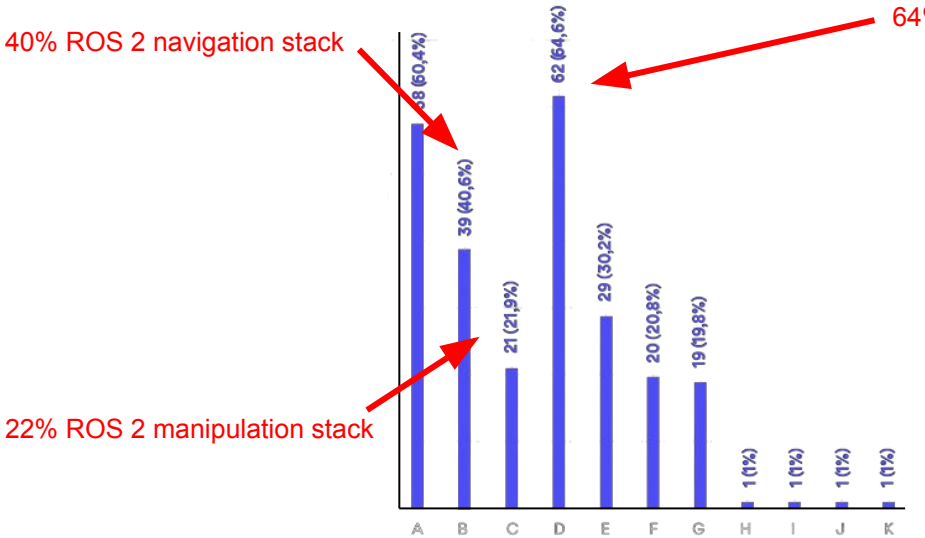


A. Gazebo/Ignition physic engines
B. ROS 2 navigation stack (navigation2)
C. ROS 2 manipulation stack (MoveIt2)
D. ROS 2 perception stack
E. ROS 2 communication middleware (DDS)
F. ROS 2 networking stack (UDP/IP/Ethernet, more deterministic network interactions)
G. ROS 2 control stack
H. All of them are important.
I. Webots physics engine
J. Image and depth data processing pipelines.
K. Lidar drivers and perception

Acceleration Robotics. (2022). Hardware Acceleration Report in Robotics. https://accelerationrobotics.com/pdf/hardware_acceleration_robotics_2022.pdf

Commonly used Tools

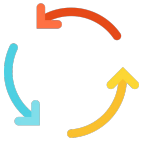
What packages/components do you think we should prioritize when it comes to hardware accelerating ROS and/or Gazebo?



A. Gazebo/Ignition physic engines
B. ROS 2 navigation stack (navigation2)
C. ROS 2 manipulation stack (MoveIt2)
D. ROS 2 perception stack
E. ROS 2 communication middleware (DDS)
F. ROS 2 networking stack (UDP/IP/Ethernet, more deterministic network interactions)
G. ROS 2 control stack
H. All of them are important.
I. Webots physics engine
J. Image and depth data processing pipelines.
K. Lidar drivers and perception

Acceleration Robotics. (2022). Hardware Acceleration Report in Robotics. https://accelerationrobotics.com/pdf/hardware_acceleration_robotics_2022.pdf

Robotics Pipeline



Robotics Pipeline

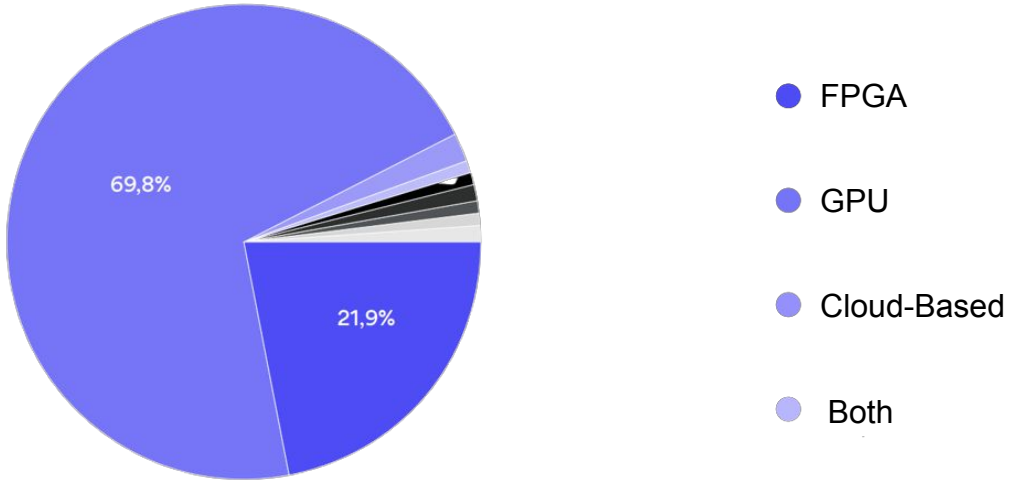
Example Kernels

Commonly used Tools

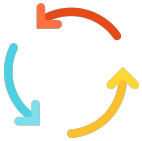
Which hardware acceleration solution are you using or planning to use?

Commonly used Tools

Which hardware acceleration solution are you using or planning to use?



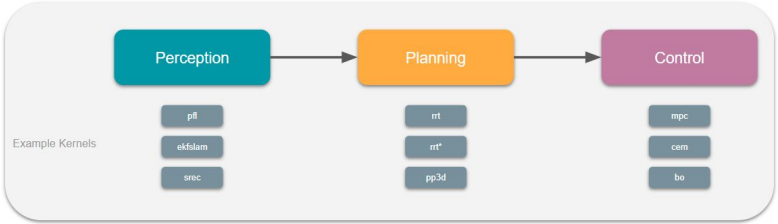
Acceleration Robotics. (2022). Hardware Acceleration Report in Robotics. https://accelerationrobotics.com/pdf/hardware_acceleration_robotics_2022.pdf



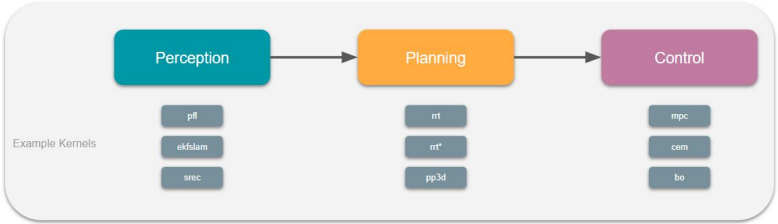
Robotics Pipeline

Example Kernels

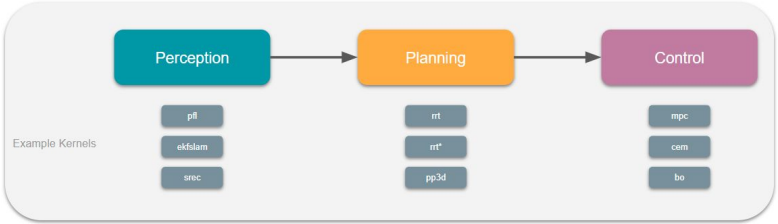
Robotics Pipeline



GPU

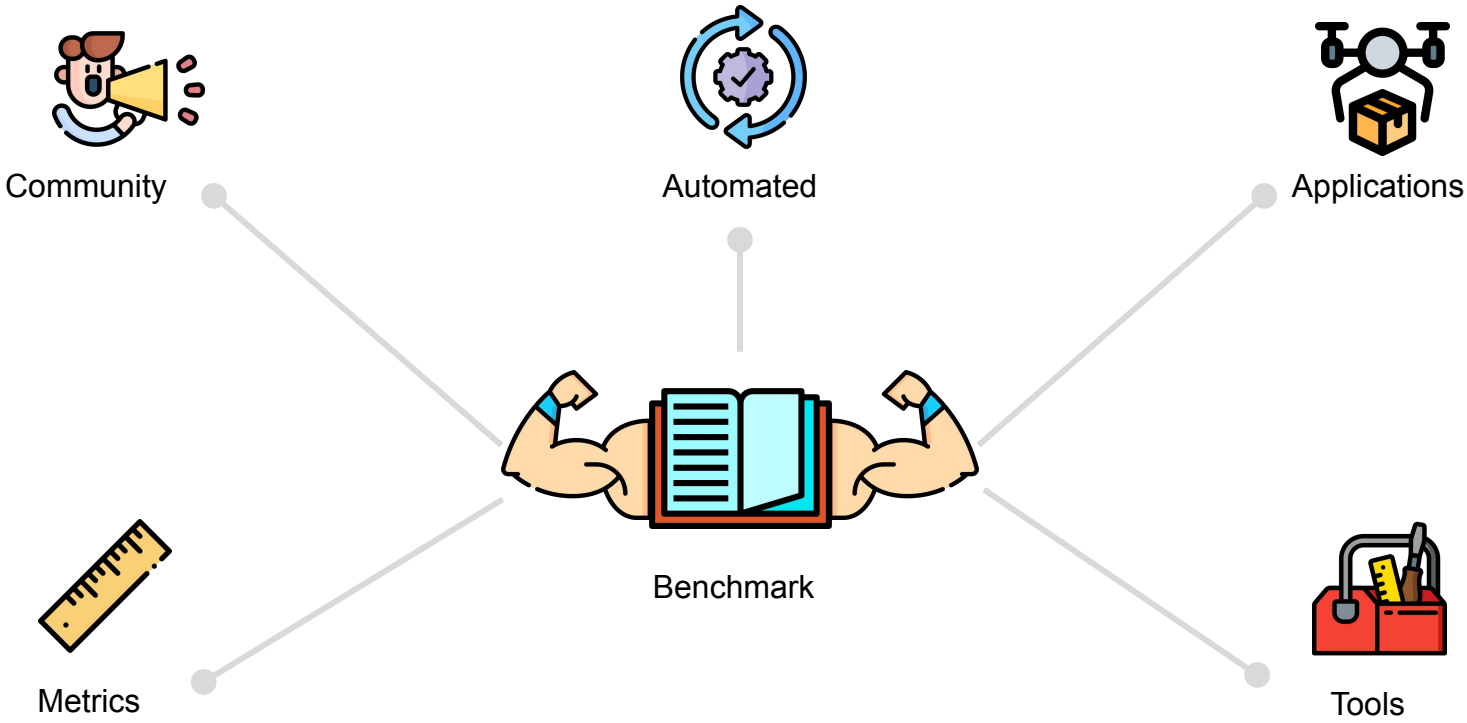


FPGA



CPU

Characteristics of a Strong Benchmark

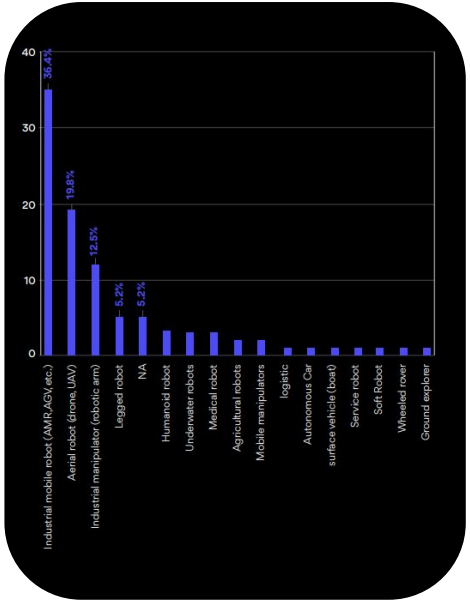


Leading Applications

What type of robot are you creating?

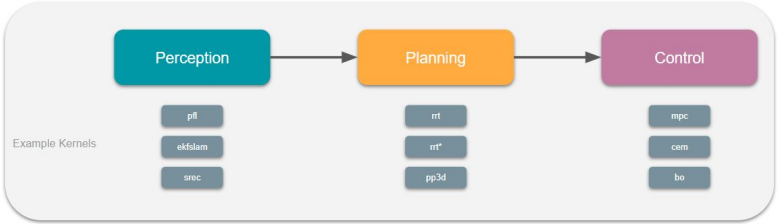
Leading Applications

What type of robot are you creating?



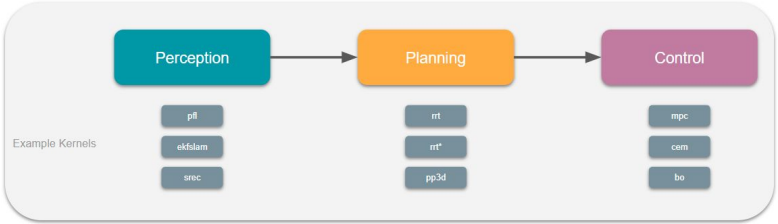
Acceleration Robotics. (2022). Hardware Acceleration Report in Robotics. https://accelerationrobotics.com/pdf/hardware_acceleration_robotics_2022.pdf

Robotics Pipeline



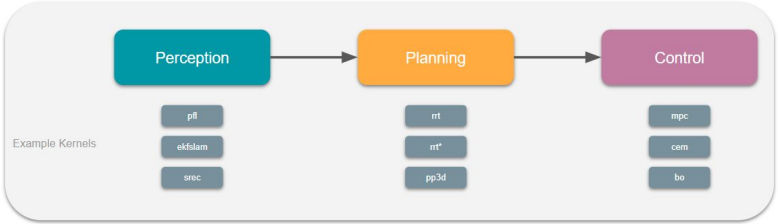
+

GPU



+

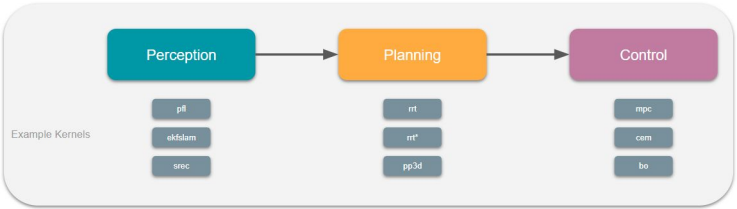
FPGA



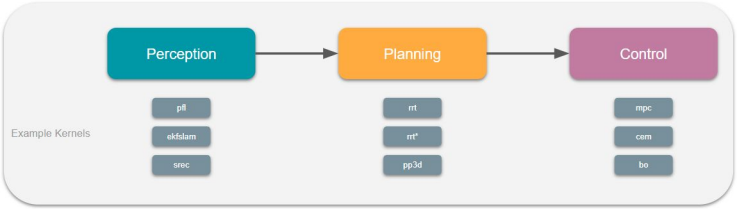
+

CPU

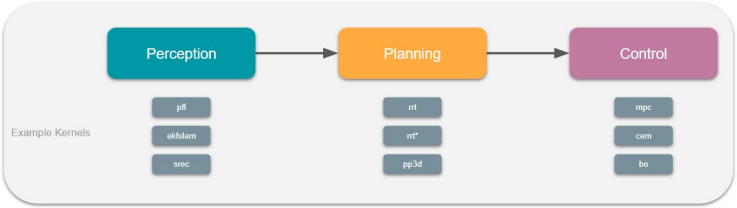
Robotics Pipeline



GPU

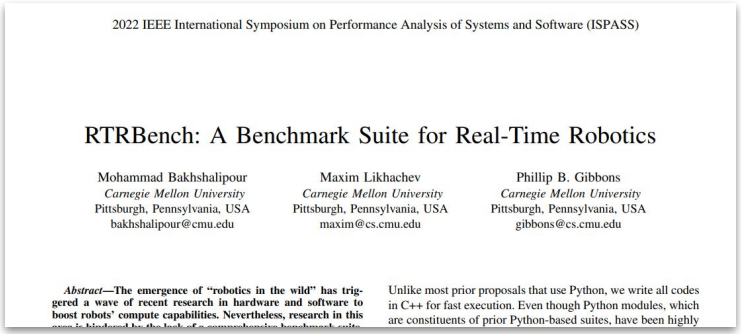


FPGA



CPU

Related Work



2022 IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS)

RTRBench: A Benchmark Suite for Real-Time Robotics

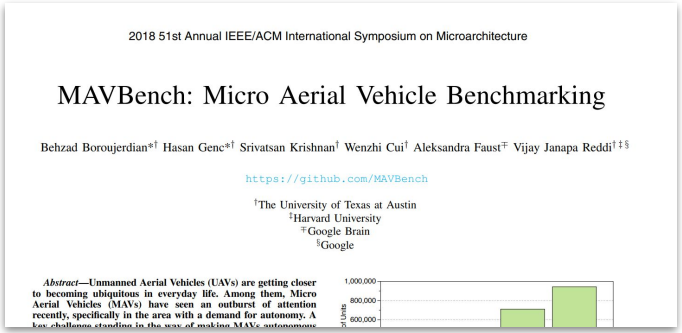
Mohammad Bakhshalipour
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA
bakhshalipour@cmu.edu

Maxim Likhachev
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA
maxim@cs.cmu.edu

Phillip B. Gibbons
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA
gibbons@cs.cmu.edu

***Abstract*—The emergence of "robotics in the wild" has triggered a wave of recent research in hardware and software to boost robots' compute capabilities. Nevertheless, research in this**

Unlike most prior proposals that use Python, we write all codes in C++ for fast execution. Even though Python modules, which are constituents of prior Python-based suites, have been highly



2018 51st Annual IEEE/ACM International Symposium on Microarchitecture

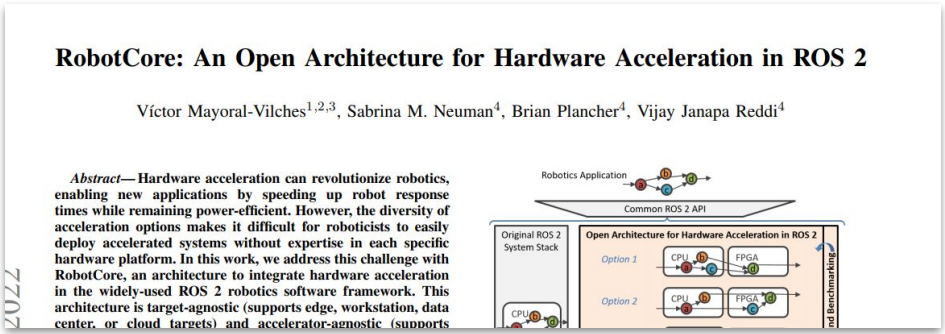
MAVBench: Micro Aerial Vehicle Benchmarking

Behzad Boroujerdian*† Hasan Genc*† Srivatsan Krishnan† Wenzhi Cui† Aleksandra Faust‡ Vijay Janapa Reddi†‡§

https://github.com/MAVBench

†The University of Texas at Austin
‡Harvard University
‡Google Brain
§Google

***Abstract*—Unmanned Aerial Vehicles (UAVs) are getting closer to becoming ubiquitous in everyday life. Among them, Micro Aerial Vehicles (MAVs) have seen an outburst of attention recently, specifically in the area with a demand for autonomy. A**



RobotCore: An Open Architecture for Hardware Acceleration in ROS 2

Víctor Mayoral-Vilches[1,2,3], Sabrina M. Neuman[4], Brian Plancher[4], Vijay Janapa Reddi[4]

***Abstract*— Hardware acceleration can revolutionize robotics, enabling new applications by speeding up robot response times while remaining power-efficient. However, the diversity of acceleration options makes it difficult for roboticists to easily deploy accelerated systems without expertise in each specific hardware platform. In this work, we address this challenge with RobotCore, an architecture to integrate hardware acceleration in the widely-used ROS 2 robotics software framework. This architecture is target-agnostic (supports edge, workstation, data center, or cloud targets) and accelerator-agnostic (supports**

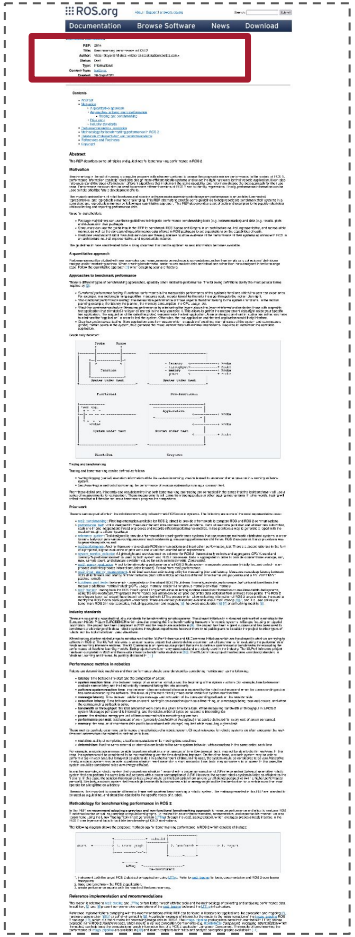
[1] Bakhshalipour, M., Likhachev, M., & Gibbons, P. B. (2022, May). RTRBench: A Benchmark Suite for Real-Time Robotics. In *2022 IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS)* (pp. 175-186). IEEE.
[2] Boroujerdian, B., Genc, H., Krishnan, S., Cui, W., Faust, A., & Reddi, V. (2018, October). Mavbench: Micro aerial vehicle benchmarking. In *2018 51st annual IEEE/ACM international symposium on microarchitecture (MICRO)* (pp. 894-907). IEEE.
[3] Mayoral-Vilches, V., Neuman, S. M., Plancher, B., & Reddi, V. J. (2022). RobotCore: An Open Architecture for Hardware Acceleration in ROS 2. *arXiv preprint arXiv:2205.03929*.

Standardization effort

[REP Index] [REP Source]

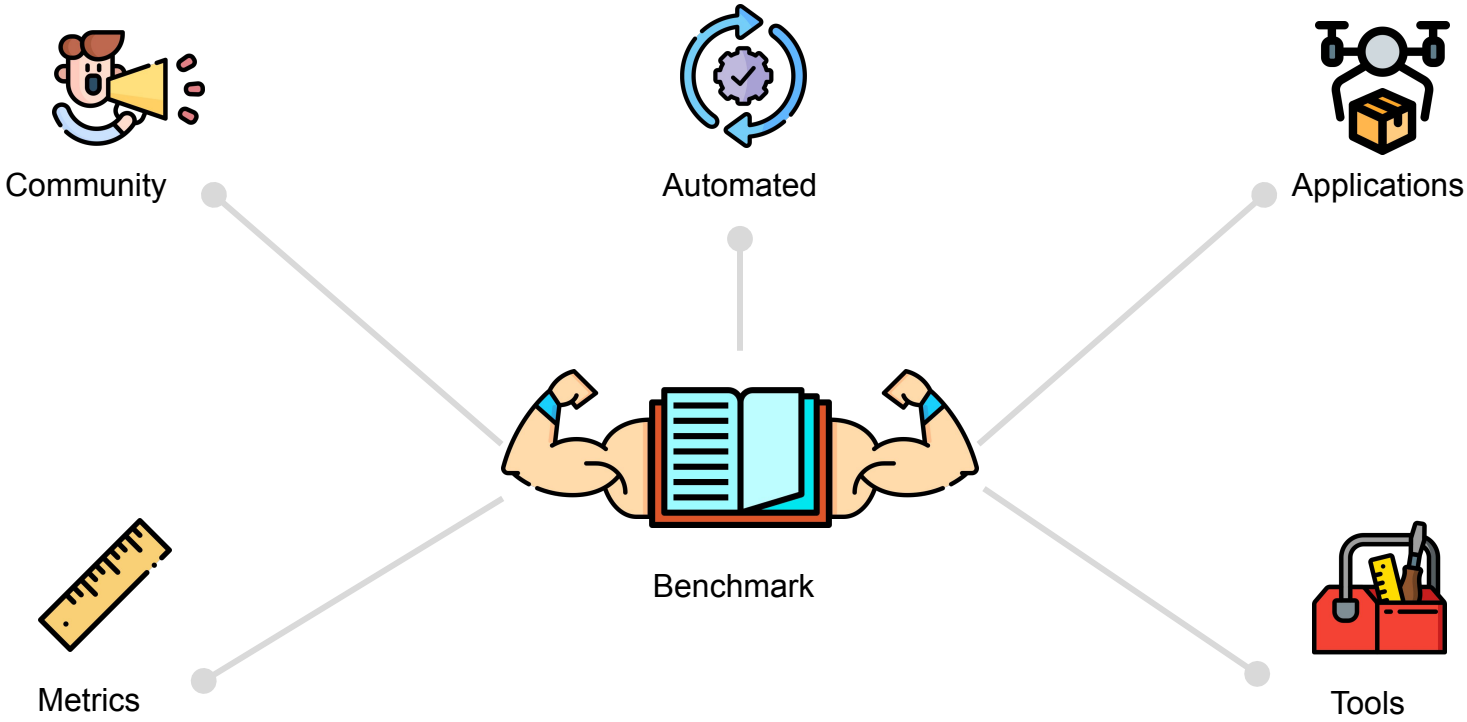
REP:	2014
Title:	Benchmarking performance in ROS 2
Author:	Víctor Mayoral-Vilches <victor at accelerationrobotics.com>
Status:	Draft
Type:	Informational
Content-Type:	text/x-rst
Created:	29-Sept-2022
Post-History:	11-Oct-2022

https://github.com/ros-infrastructure/rep/pull/364



[1] REP-2014 Pull Request. Benchmarking performance in ROS 2. https://github.com/ros-infrastructure/rep/pull/364

Characteristics of a Strong Benchmark



Example Leaderboard

Live @ Mobile AI CVPR 2022 Workshop ● Tutorials from MediaTek, Qualcomm, Intel, Huawei, OPPO, Google and AI Benchmark

Phones (V4) | Mobile SoCs | IoT

Performance Ranking

Desktop GPUs and CPUs

View Detailed Results

Model	SoC	RAM	Year	Android	Updated	Lib	CPU-Q Score	CPU-F Score	INT8 NNAPI 1.1	INT8 NNAPI 1.3	INT8 Accuracy	FP16 NNAPI 1.1	FP16 NNAPI 1.3	FP16 Accuracy	INT8 Parallel	FP16 Parallel	INT8 NLP	FP16 NLP	INT8 Memory	FP16 Memory	AI Score, K
Asus ROG Phone 6	Snapdragon 8+ Gen 1	12GB	2022	12L	10.22	qhqh	18454	20721	168080	445042	82.9	72053	113016	91.2	53622	25169	33800	59092	1600	1600	1447
Xiaomi 12S Pro	Snapdragon 8+ Gen 1	8GB	2022	12	10.22	qhqh	16908	18509	155472	442784	82.9	70415	112104	91.2	60109	24606	33546	58912	1700	1600	1431
Lenovo Legion Y70	Snapdragon 8+ Gen 1	12GB	2022	12L	10.22	qhqh	13517	16071	161533	442559	82.9	70963	112229	91.2	62826	23787	33775	58762	1600	1600	1429
Xiaomi 12S Ultra	Snapdragon 8+ Gen 1	12GB	2022	12	10.22	qhqh	15962	17673	153527	440010	82.9	68273	112949	91.2	54779	25241	32692	58476	1700	1600	1411
Xiaomi 12S	Snapdragon 8+ Gen 1	8GB	2022	12	10.22	qhqh	16308	17733	150957	440512	82.9	70078	112127	91.2	53657	24804	32326	58347	1600	1600	1399
vivo iQOO 10 Pro	Snapdragon 8+ Gen 1	12GB	2022	12L	10.22	qhqh	12116	15358	161777	434986	82.9	71695	111032	91.2	55224	24386	33077	58819	1500	1400	1387

Ignatov, A., Timofte, R., Kulik, A., Yang, S., Wang, K., Baum, F., ... & Van Gool, L. (2019, October). Ai benchmark: All about deep learning on smartphones in 2019. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)* (pp. 3617-3635). IEEE.

Thank you!

Contact us:

Prof. Vijay Janapa Reddi

vj@eecs.harvard.edu

Jason Jabbour

jasonjabbour@g.harvard.edu

Víctor Mayoral-Vilches

victor@accelerationrobotics.com