arm

On-device Machine Learning

Delivering endpoint AI for mobile devices

Solution Brief

As technology enters a new wave of innovation driven by artificial intelligence (AI), machine learning (ML) sits at the forefront of its development. ML uses statistical techniques to give devices the ability to "learn" through receiving data, without needing to be explicitly programmed. In the coming years, ML will be rapidly introduced into more devices worldwide, achieving a new level of 'intelligence.'

Four fundamentals of on-device ML

- Improved performance endpoint
- Decreased latency
- Fewer security concerns
- Scalability

The fundamentals

The power and cost required to shift massive amounts of data to and from the cloud are prohibitive and can produce a noticeable lag or delay – something that time-critical applications simply cannot tolerate. Therefore, devices need to have the capability and flexibility to perform ML tasks at the endpoint – on the device – rather than in the cloud. The latency caused by constantly interacting with the cloud means that tasks on a mobile device perform at slower speeds. Moreover, sending data back and forth from the cloud creates a system more vulnerable to security threats.

Scalability is also crucial for ML technologies, with architecture needing to be scaled up or down based on the type of device and its required performance. Being able to offer ML on a variety of devices, including mobile phones, DTVs, sensors and smart speakers, and across the performance curve, is important in the future development of technologies.

ML on Arm-powered devices

ML is a vital area of innovation, particularly across mobile. Arm's suite of premium IP, which features the latest **Arm Cortex-A processors**, **Arm Mali GPU processors**, and **Arm Ethos-N processors** for the next generation mobile devices, further optimizes ML at the edge. It also feeds perfectly into Arm's heterogenous AI platform.

Why ML on Arm

- Better endpoint AI performance
- Industry-leading power efficiency
- Technology and solutions from a trusted technology company
- Greater scalability to target the specific requirements of their SoC

The Arm AI platform

The Arm AI platform represents a suite of Arm products that gives device-makers all the hardware and software choices they need for on-device ML. In addition to the latest CPUs and GPUs outlined previously, the platform consists of a range of hardware and software products including Arm NN, the Arm Compute Library, and CMSIS-NN.

The Arm AI platform is particularly suitable for the mobile market but is easily scaled across multiple devices. It also has the flexibility to enable a new class of ML-equipped devices with advanced compute capabilities.

Cortex-A CPUs

- Deliver compute performance improvements for ML at the edge, enabling improved, more responsive and more secure ML user experiences on mobile devices.

Mali GPUs

- Deliver a high GPU performance, providing greater flexibility for ML at the edge.

Arm Ethos NPUs

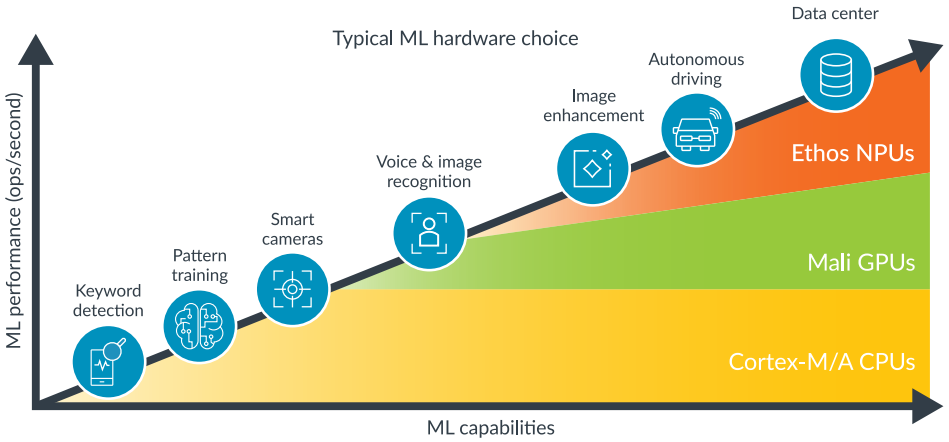
- Designed specifically for inference at the edge.
- Provide a massive efficiency uplift from CPUs, GPUs, DSPs and accelerators.
- Provide unmatched performance in thermal- and cost-constrained environments.

Arm NN

- Software framework that enables the efficient translation of existing neural networks to support ML workloads across all Arm programmable IP.
- Provides support for Cortex-A CPUs, Mali GPUs and the Ethos NPU.
- Supports leading NN frameworks, including TensorFlow, Caffe, Android NNAPI and MXNet Graph and kernel optimizations for each IP type.
- Allows developers to get the highest performance from ML applications by being able to fully utilize underlying Arm hardware capabilities and performance.
- Available free of charge, under a permissive MIT open-source license.

Flexible, Scalable AI Solutions

Only Arm enables AI everywhere



Benefits of Arm ML solution for mobile devices

Arm's ML solutions provide improved performance for devices at the edge, while also offering the flexibility for partners to use the IP and processors in future devices as ML innovation continues to develop. For more information visit www.arm.com/solutions/artificial-intelligence.