



The Dream of Edge AI

Author: Kris Ardis

At this point, we should have had flying cars. And robot butlers. And with some bad luck, sentient robots that decide to rise up against us before we can cause the apocalypse. While we don't have that, it is clear that artificial intelligence (AI) technology has made its way into our world. Every time you ask Alexa to do something, machine learning technology is figuring out what you said and trying to make the best determination on what you wanted it to do. Every time Netflix or Amazon recommend that "next movie" or "next purchase" to you, it is based on sophisticated machine learning algorithms that give you compelling recommendations that are far more enticing than sales promotions of the past. And while we might not all have self-driving cars, we're all keenly aware of the developments in that space and the potential that autonomous navigation can offer.

AI technology carries a great promise – the idea that machines can make decisions based on the world around them, processing information like a human might (or in a manner superior to what a human would do). But if you think about the examples above, the AI promise here is only being fulfilled by "big machines"—things that don't have power, size, or cost constraints, or to put it another way—they can get hot, have line power, are big, and are expensive. Alexa and Netflix rely on big, power hungry servers in the cloud to figure out your intent. While self-driving cars are likely to rely on batteries, their energy capacity is enormous, considering those batteries must turn the wheels and steer, which are big energy expenses compared to even the most expensive AI decisions.

While the promise of AI is great, "little machines" are being left behind. Devices that are powered by smaller batteries or have cost and size constraints are unable to participate in the idea that machines

can see and hear. Today, these little machines can only make use of simple AI technology: perhaps listening for a single keyword or analyzing low-dimensional signals like photoplethysmography (PPG) from a heart rate.

What if little machines could see and hear?

But is there value in small machines being able to see and hear? It is hard to think about things like a doorbell camera taking advantage of technologies like autonomous driving or natural language processing, but there is an opportunity for less complex, less processing intensive AI computations such as vocabulary recognition, voice recognition, and image analysis.

- Doorbell cameras and consumer security cameras often trigger on uninteresting events, such as motion of plants caused by wind, drastic light changes caused by clouds, or even events such as dogs or cats running in front. This can result in false triggers, causing the homeowner to begin to ignore the events. In addition, if the homeowner is travelling in a different part of the world, they are probably sleeping while their camera is alarming to changes in lighting caused by sunrise, clouds, and sunset. A smarter camera could trigger on more specific events, such as a human being in the frame of reference.
- Door locks or other access points can use facial identification or even speech recognition to grant access to authorized personnel, forgoing the need for keys or badges in some cases.
- Lots of cameras want to trigger on certain events: for instance, trail cameras might want to trigger on the presence of a deer in the frame, security cameras might want to trigger on a person in the frame or a noise like a door opening or footsteps, and a personal camera might want to trigger with a spoken command.
- Large vocabulary commands can be useful in many applications: while there are plenty of “Hey Alexa” solutions, if you start to think about a vocabulary of 20 or more words, you can find use in industrial equipment, home automation, cooking appliances, and plenty of other devices to simplify the human interaction.

These examples only scratch the surface: the idea of allowing small machines to see, hear, and solve problems that in the past would require human intervention is a powerful one and we continue to find creative new use cases every day.

What are the challenges to enabling little machines to see and hear?

So, if AI could be so valuable to little machines, why don't we have it yet? The answer is computational horsepower. AI inferences are the result of the computation of a neural network model. Think of a neural network model as a rough approximation for how your brain would process a picture or a sound, breaking it into very small pieces and then recognizing the pattern when those small pieces are put together. The workhorse model of modern vision problems is the convolutional neural network (CNN). These kinds of models are excellent at image analysis and are very useful in audio analysis as well. The challenge is that these models take millions or billions of mathematical computations. Traditionally, these applications have a difficult choice to make for implementation:

- Use an inexpensive and low-powered microcontroller solution. While the average power consumption may be low, the CNN can take seconds to compute, meaning the AI inference is not real time, and meaning it consumes considerable battery power.

- Buy an expensive and high-powered processor that can complete those mathematical operations in the required latency. These processors are typically large and require lots of external components including heat sinks or similar cooling components. However, they execute AI inferences very quickly.
- Don't implement. The low power microcontroller solution will be too slow to be useful, and the high-powered processor approach will break cost, size, and power budgets.

What is needed is an embedded AI solution built from the ground up to minimize the energy consumption of a convolutional neural network computation. AI inferences need to execute at orders of magnitude with less energy than conventional microcontroller or processor solutions, and without the assistance of external components such as memories, which consume energy, size, and cost. If an AI inferencing solution could practically eliminate the energy penalty of machine vision, then even the smallest devices could see and recognize things happening in the world around them.

Lucky for us, we are at the beginning of this – a revolution of the “little machines”. Products are now available to nearly eliminate the energy cost of AI inferences and enable battery powered machine vision. [Find out more about a microcontroller built to execute AI inferences while spending only microjoules of energy.](#)