Why is Machine Learning on Embedded Systems so Important?

petewarden@google.com

Slides at goo.gle/emb-ml-importance

We can capture much more data than we can send to the cloud

Is this true?

- What are the limits on sending data to the cloud?
- How much sensor data can we capture on the edge?
- What opportunities are we missing without this data?
- How does machine learning help?

Slides at goo.gle/emb-ml-importance

Limits on Sending Data

Why is it hard to send data to the cloud?

Phones and mains-powered devices:

- Lack of geographic coverage
- Electronically-noisy environments
- Transmission costs
- Cloud processing costs
- Latency

Slides at goo.gle/emb-ml-importance

Why is it hard to send data to the cloud?

In ideal circumstances, you might get broadband speeds from your phone.

Most of the time, in most places, that won't be true, and is unlikely to change in the medium term.

The most successful "near edge" devices with high bandwidth expectations are security cameras, so it is possible. Even these rely on a lot of client-side engineering (compression, deltas, buffering, latency tolerance) to work, and charge for storage.

Slides at goo.gle/emb-ml-importance

Google Translate and Word Lens

One of my favorite features in a Google app is the live camera translation in Translate. It's great for menus and signs.

Originally, this worked by sending a single photo to the cloud and doing server-side processing.

WordLens was a startup offering real-time overlaid translation on a phone.

The metrics were worse, but the user experience was much better. It felt more accurate.

So, getting data from phones is hard...

… but it’s even worse for other devices!

If a device isn’t mains-powered or recharged every night, there are severe constraints on the amount of energy it can use.

How much power can you get from an AA?

There are around 10,000 Joules in an AA battery.

There are 31,536,000 seconds in a year.

So, if you rely on one AA battery, in theory you've got an average 0.000317 watts of continuous power, or 317 microwatts.

In practice it's a lot more complex with battery characteristics, but a rule of thumb is that three AAs might give you about one milliwatt of power for a year, if you're lucky.

Slides at goo.gle/emb-ml-importance

https://commons.wikimedia.org/wiki/File:02_-_Single_Energizer_Battery.jpg

What about energy harvesting?

Method	Microwatts per cm²
Outdoor Solar	10,000
Industrial Temperature	10,000
Industrial Vibration	100
Human Temperature	25
Indoor Solar	10
Human Vibration	4
GSM Radio	0.1
Wifi Radio	0.001

See goo.gle/energy-data

In most environments, there's less than 1,000 microwatts, or one milliwatt, available from a realistic form factor.

What does that all mean?

The bottomline is that any unattended device without a mains connection will be lucky to have an average of one milliwatt of power.

Why does this matter for data transmission?

Current radio technologies all use a lot more than one milliwatt when they're active.

Even BLE is often around 100 milliwatts, with Wifi, cell, or satellite technologies requiring even more.

That means we can't power continuous transmission of data to the cloud, or even a reasonable duty cycle, from these kinds of devices.

Slides at goo.gle/emb-ml-importance

Limits on capturing data

How much data can we capture in one mW?

Himax HM01B0 image sensor: 30 FPS at 320x320 = 3MB per second

Vesper VM1010: 16,000 Hz = 32 kilobytes per second (for 300 mW)

Bosch BMI270 IMU: 6,400 Hz = 13 kilobytes per second

Slides at goo.gle/emb-ml-importance

How much data can we process in one mW?

Ambiq Cortex M4 at 48MHz clock rate means we can handles tens of megabytes a second, even with conservative estimates.

Slides at goo.gle/emb-ml-importance

The Great Data Bottleneck

Slides at goo.gle/emb-ml-importance

We can capture much more data than we can send to the cloud

Why does this mismatch matter?

Existing devices can’t reach most of the world

Current computing devices are joined at the hip to people. We have to do one of the following:

- Run a mains wire to a location and plug it in.
- Carry it around and charge it every night.
- Change its batteries every few days.

Slides at goo.gle/emb-ml-importance

What opportunities are we missing?

- Peel-and-stick sensors for homes, offices, and factories
- Agricultural disease and pest monitoring
- Medical devices
- Wildlife and ecosystem protection

Slides at goo.gle/emb-ml-importance

Peel and stick sensors?

This is the idea that you can deploy a sensor-based computing device by simply peeling off a backing and placing it where you want it. At the moment, installing almost any home, office, or factory device requires:

- Entering a Wifi or other connection information through an app.
- Configuring location information.

The vision is that we'll be able to drastically reduce the human labor needed to deploy sensors, to the point where we can scatter and forget them.
https://www.energy.gov/eere/buildings/articles/low-cost-peel-and-stick-wireless-sensors-prototype-development-field-testing

How does ML help?

How does machine learning help?

Advances in machine learning since 2010 have centered on the application of neural networks and deep learning.

These have proven to be very effective in taking noisy sensor data (images, audio, accelerometers) and robustly spotting patterns (objects, speech, noises, gestures).

Slides at goo.gle/emb-ml-importance

Isn’t ML something for data centers?

Thousands of applications and billions of users already run TensorFlow Lite on their iOS and Android phones!

Useful models (for example wake word detection) can fit in less than 30KB.

They can take less than a million arithmetic operations to execute.

It’s possible to run them without floating-point support.

Slides at goo.gle/emb-ml-importance

ML on Embedded

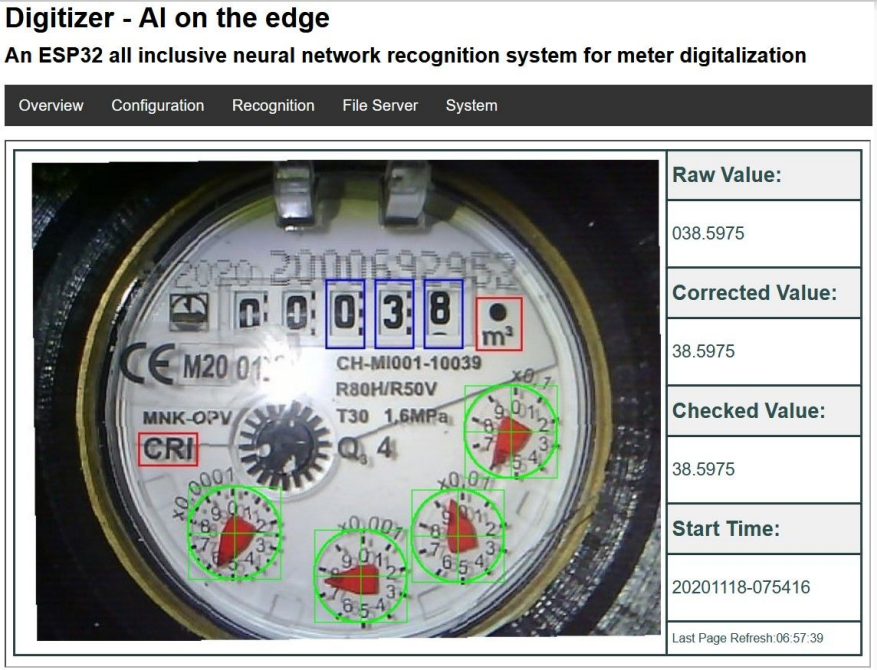
Can take those streams of sensor data and turn them into actionable information:

- Is there a crop pest in this field?
- Is there a person on the subway track?
- Did someone give a voice command?

ML fills the gap between the sensor and radio capabilities. The radio can be turned on when something important happens.

Slides at goo.gle/emb-ml-importance

Reading water meters



https://petewarden.com/2021/02/28/how-screen-scraping-and-tinyml-can-turn-any-dial-into-an-api/

What are some other applications?

- Voice interfaces on every consumer item
- Wearable devices you don’t need to charge daily
- How many people are in the conference room?
- Agricultural and wildlife monitoring
- Sensors embedded into buildings

Slides at goo.gle/emb-ml-importance

Thanks!

Am I right? Questions?

petewarden@google.com
twitter.com/petewarden

Slides at goo.gle/emb-ml-importance