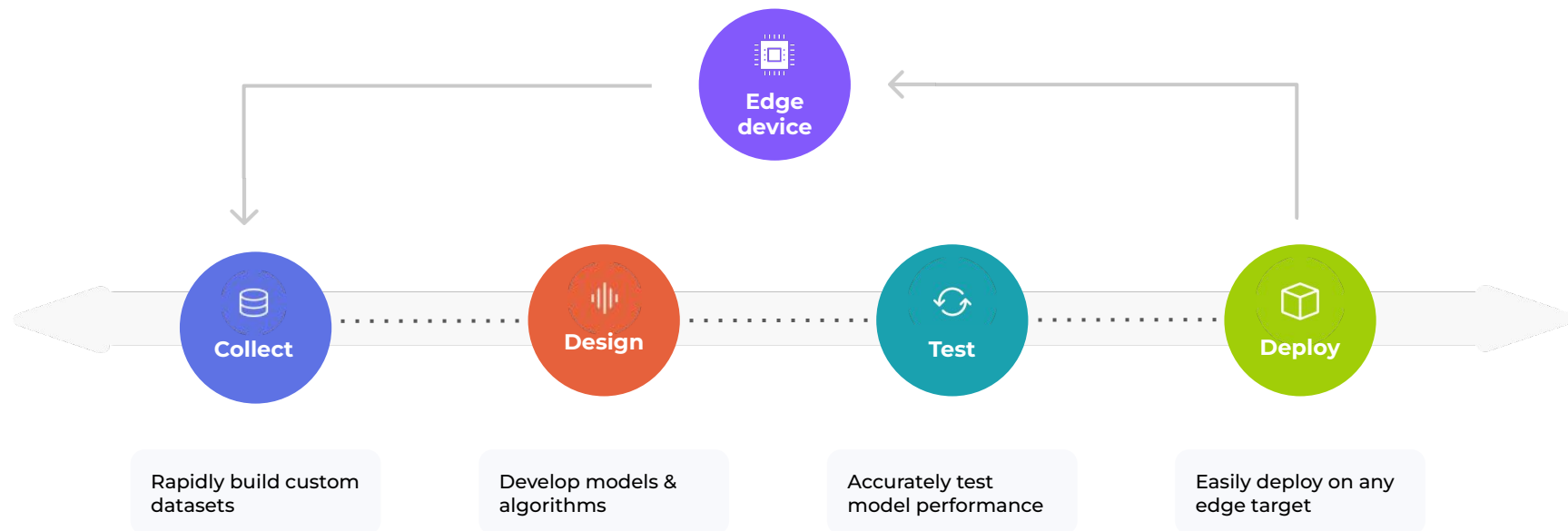


# Making the impossible, possible

Aurelien Lequertier, VP Solutions

# Develop edge ML applications with Edge Impulse

The infrastructure and integrations your data science and ML teams need.



# Empower all enterprises and developers with ML

**40,000+**

Developers

**75,000+**

Projects

**1,000+**

Enterprises

TRUSTED BY LEADING ENTERPRISES

ŌURA

ADVANTECH

 poly

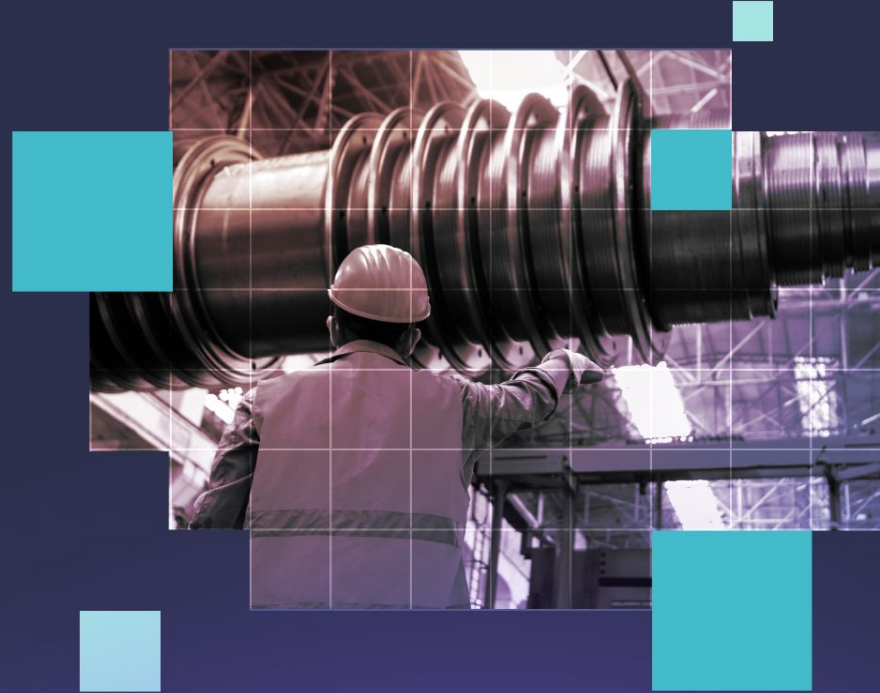


SONY

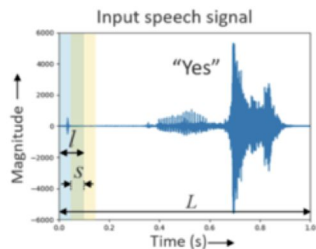
# Any sensor, any data, any use case

	Ultra low power	Low-end MCU	High-end MCU	NPU	MPU	GPU
Memory	Anomaly detection 10kB	Sensor fusion classification 18kB	Audio classification 50kB	Image classification 256kB	Object detection complex voice processing 1MB	Video classification 1GB+
Sensor	✓	✓	✓	✓	✓	✓
Audio	✓	✓	✓	✓	✓	✓
Image			✓	✓	✓	✓
Video					✓	✓

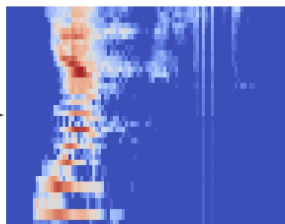
# Few-Shot Keyword Spotting



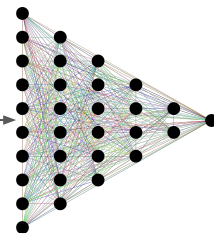
# How Keyword Spotting works?



Collecting data



Features extraction  
(MFCC)



Neural Network

# Some typical Keyword Spotting model

- Extracted features fed into NN
- 1D or 2D Convolutional network

Mel Frequency Cepstral Coefficients	
Number of coefficients	<input type="text" value="13"/>
Frame length	<input type="text" value="0.02"/>
Frame stride	<input type="text" value="0.02"/>
Filter number	<input type="text" value="32"/>
FFT length	<input type="text" value="256"/>
Normalization window size	<input type="text" value="101"/>



# Few-Shot Keyword Spotting

## Problem

Keyword spotting requires giant, diverse datasets which are difficult to source



# Few-Shot Keyword Spotting

~560 samples,  
class balanced

Last training performance (validation set)



ACCURACY  
75.2%



LOSS  
1.11

Confusion matrix (validation set)

	NO	NOISE	UNKNOWN	YES
NO	68.2%	0%	18.2%	13.6%
NOISE	3.7%	85.2%	11.1%	0%
UNKNOWN	15.6%	9.4%	56.3%	18.8%
YES	3.1%	6.3%	0%	90.6%
F1 SCORE	0.68	0.84	0.63	0.83

# Few-Shot Keyword Spotting

~10.8k samples,  
class balanced

**19x the data for 12%  
better accuracy**

Last training performance (validation set)



ACCURACY

87.3%



LOSS

0.41

Confusion matrix (validation set)

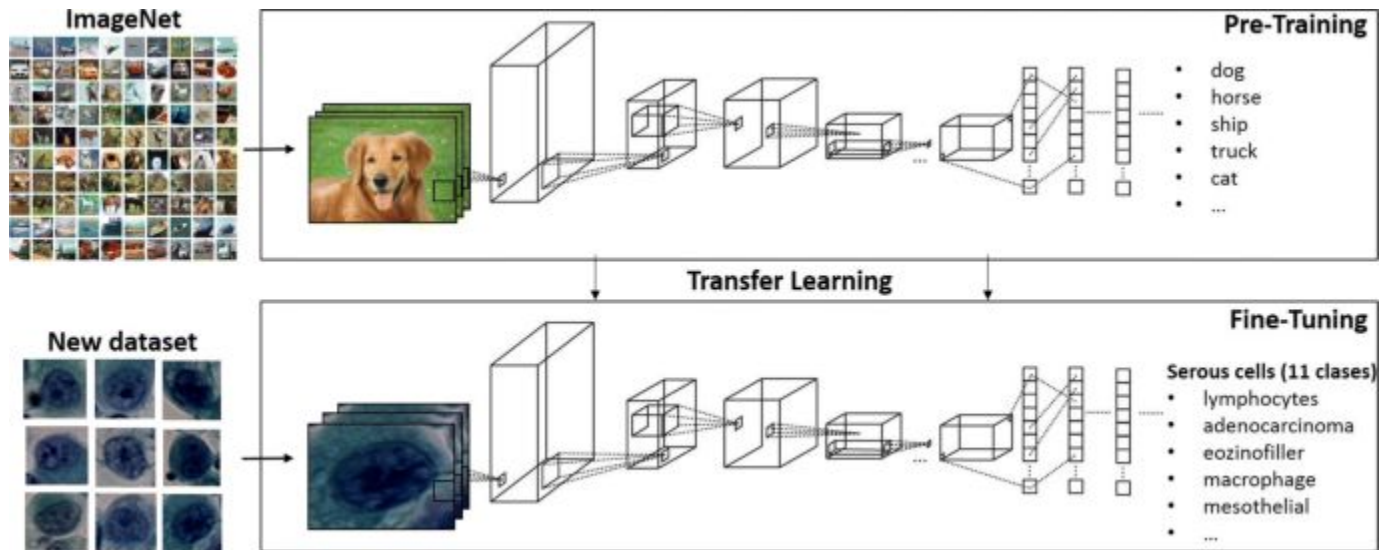
	NOISE	OFF	ON	UNKNOWN
NOISE	95.4%	0.2%	0.4%	4%
OFF	2.0%	87.1%	5.6%	5.4%
ON	4.2%	6.1%	86.4%	3.3%
UNKNOWN	10.3%	4.9%	4.5%	80.3%
F1 SCORE	0.90	0.88	0.88	0.83

# Few-Shot Keyword Spotting

## Solution

Transfer learning, just like we do with images, but for keyword spotting

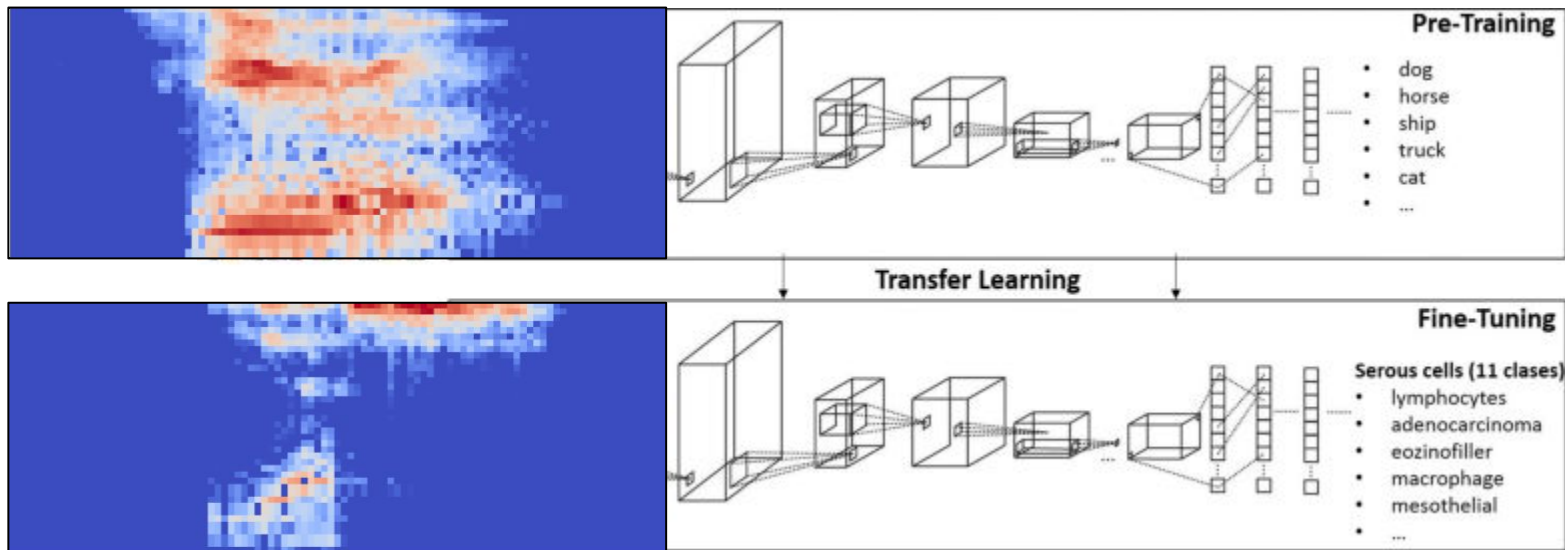
# Few-Shot Keyword Spotting



# Few-Shot Keyword Spotting

Millions of utterances in many languages

<https://www.seas.harvard.edu/news/2021/12/voice-technology-rest-world>



A few examples of specific utterances

# Few-Shot Keyword Spotting

Model trained with ~20 occurrences for each keyword

Model

Model version: ?

Unoptimized (float32) ▾

Last training performance (validation set)



ACCURACY

97.4%



LOSS

0.10

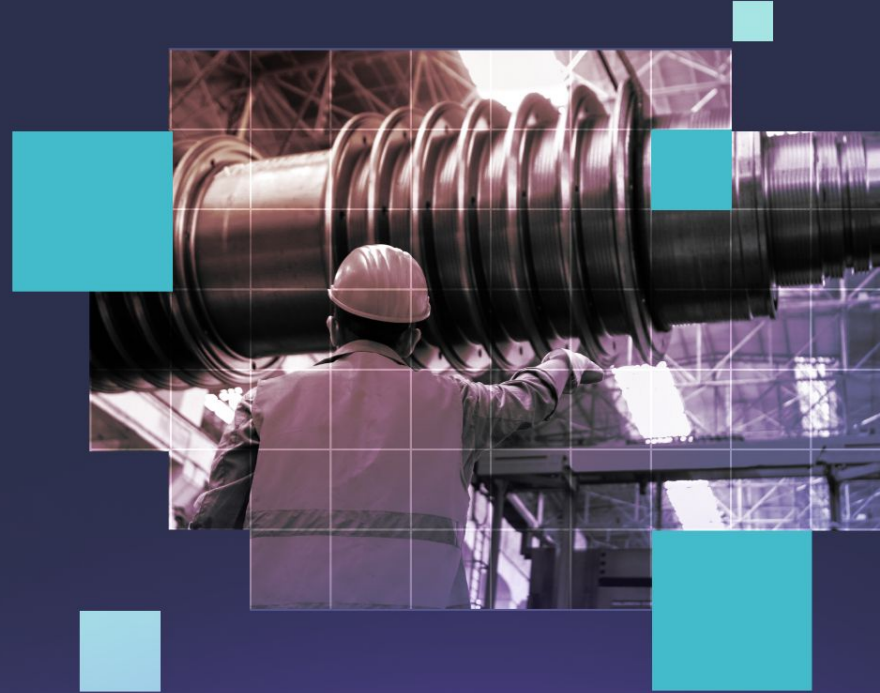
Confusion matrix (validation set)

	GREENLIGHT	NOISE	OK_EDGE	REDLIGHT	UNKNOWN
GREENLIGHT	100%	0%	0%	0%	0%
NOISE	0%	97.5%	0%	0%	2.5%
OK_EDGE	0%	0%	100%	0%	0%
REDLIGHT	0%	0%	0%	100%	0%
UNKNOWN	0%	1.2%	0%	2.4%	96.3%
F1 SCORE	1.00	0.98	1.00	0.91	0.97

# Demo

The image shows a screenshot of the Edge Impulse web dashboard. The interface features a purple header with the 'EDGE IMPULSE' logo and navigation links. A left sidebar contains a menu with options like 'Dashboard', 'Projects', 'Models', 'Hardware', 'API', 'Settings', and 'Help'. The main content area is divided into several panels: a 'Training Job' configuration panel with a green 'Train' button, a 'Model Performance' panel with a map of the United States showing data points, and a 'Model Deployment' panel with status indicators. A circular inset in the bottom-left corner shows a man with short brown hair and a beard, wearing a black t-shirt with the 'EDGE' logo, speaking into a microphone.

# FOMO (Faster Objects, More Objects)





# What is Computer Vision?

Image classification



```
2.570694 fps
Predictions at [x=0,y=0,w=240,h=240]
lamp = 0.000000
plant = 0.945313
unknown = 0.054688
2.570694 fps
*****
Predictions at [x=0,y=0,w=240,h=240]
lamp = 0.000000
```

Object detection



# Object Detection

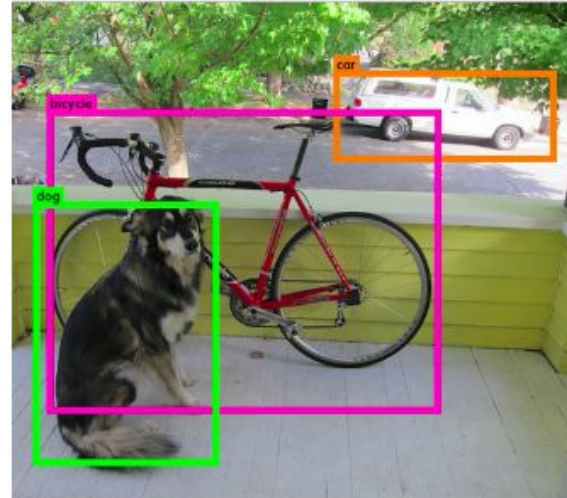
## Problem

Traditional object detection models are poorly suited to MCUs

# Object Detection



Multiple Bounding Boxes



Final Bounding Boxes

*Source: <https://pjreddie.com/darknet/yolov1/>*

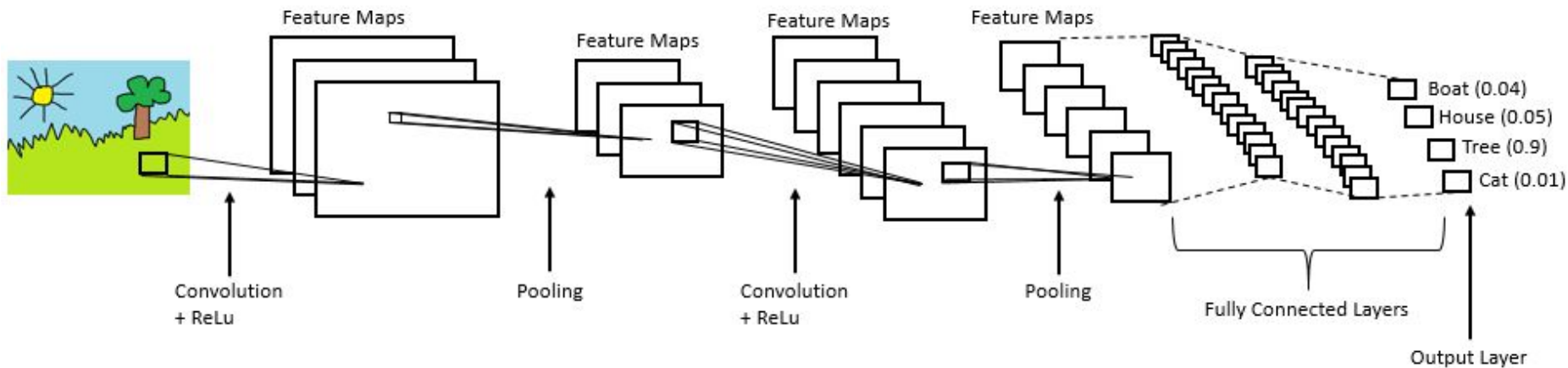


# Object Detection

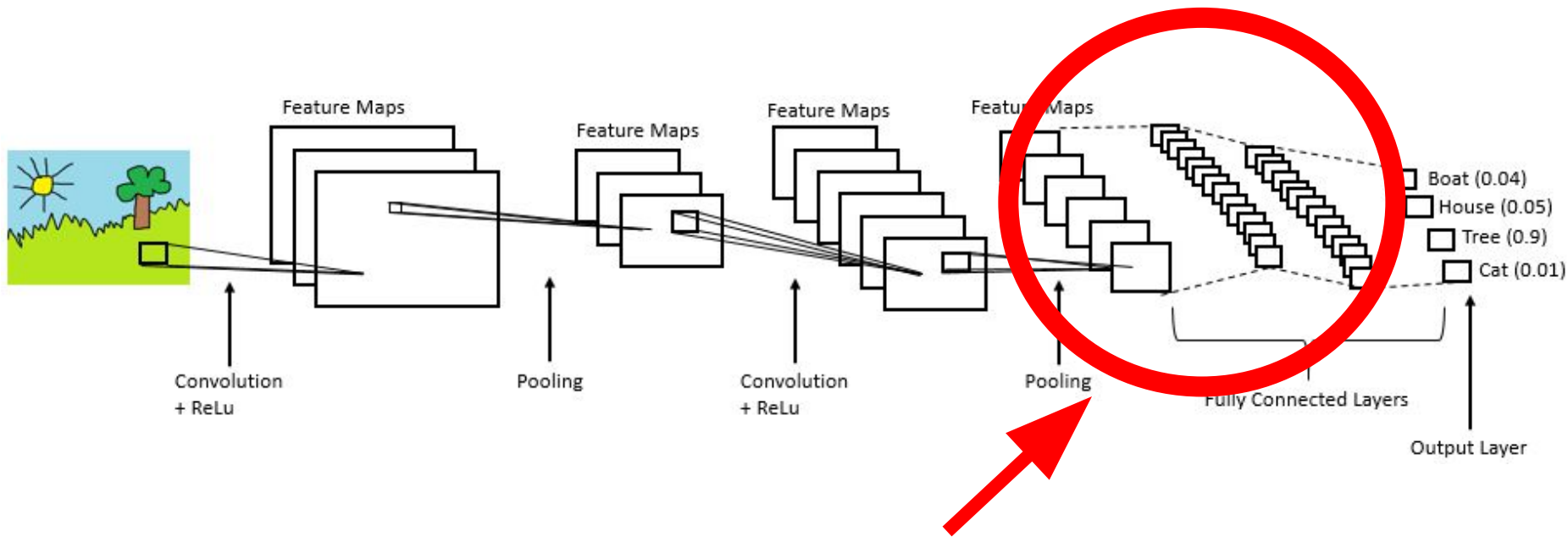
## Solution

Simplify the model so that it is smaller and better suited to the problems in scope

# Object Detection



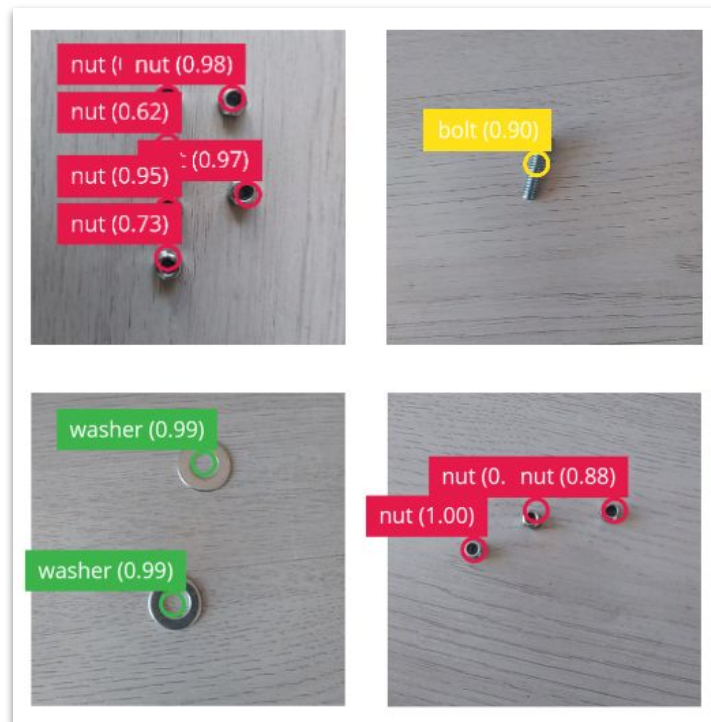
# Object Detection



Replace with single per-region class probability map

# Introducing FOMO: Faster objects, more objects

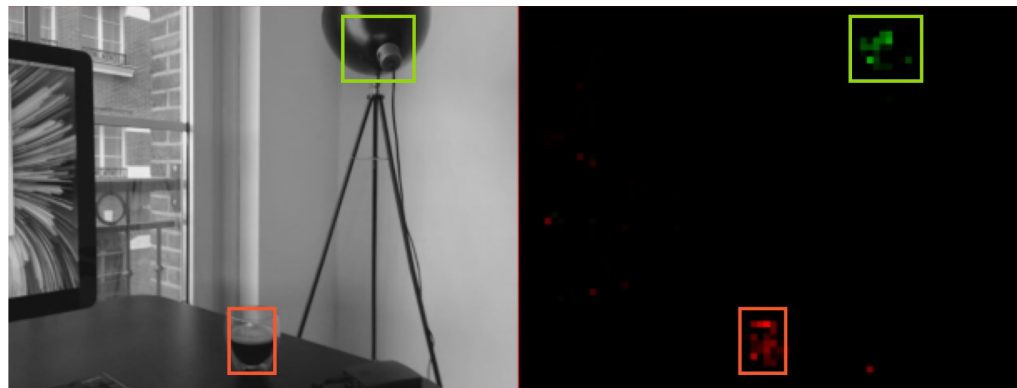
- Object detection on MCUs
- Based on MobileNet architecture
- Ultra fast: 60 fps on RPi class, 30 fps on Cortex M7 (Arduino Nicla Vision), 10 fps on Cortex M4F
- Better at detecting smaller and more numerous objects
- Capable of segmentation and counting objects



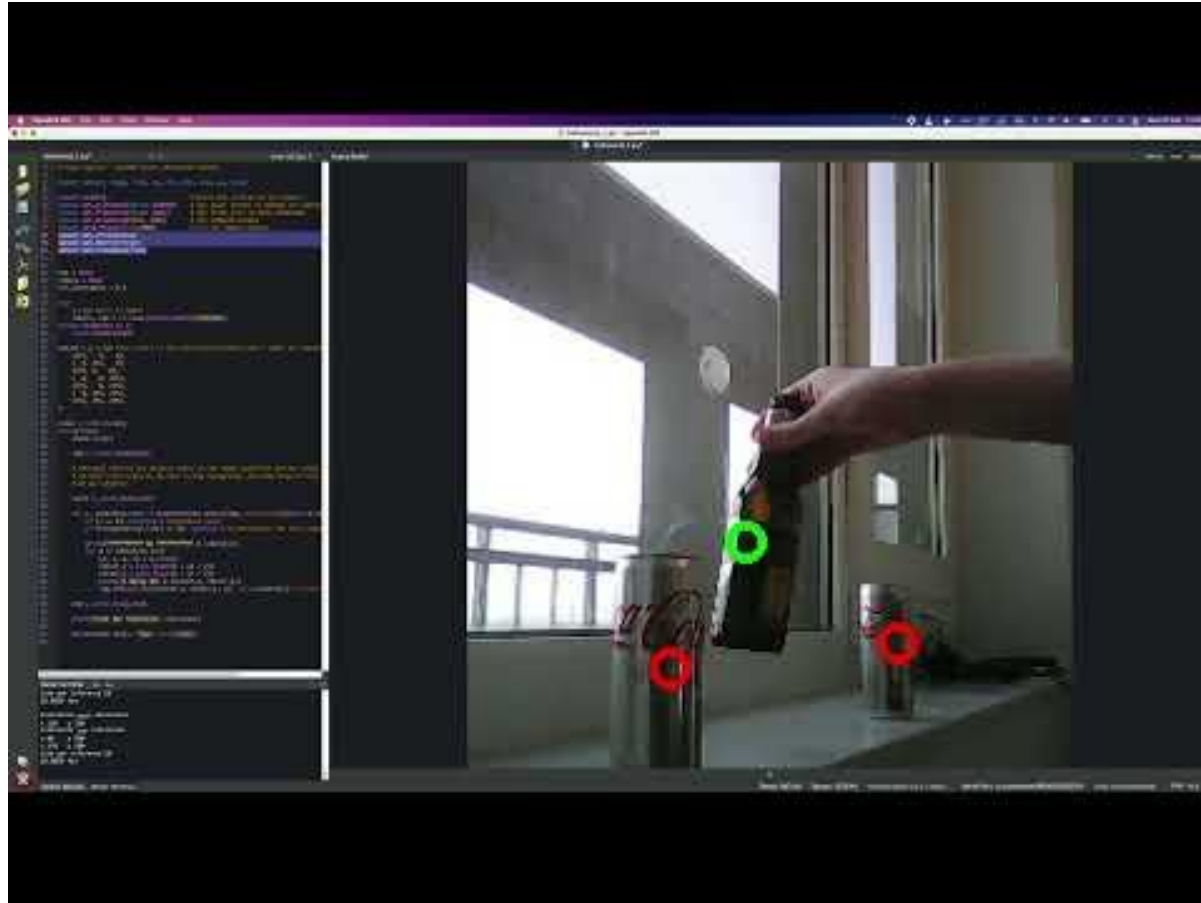


# Introducing FOMO: Faster objects, more objects

- Heatmap of objects location
- Training on centroids



# Demo



# To go further

Create your free account on [edgeimpulse.com](https://edgeimpulse.com)

Login

Get started

Build your first keyword spotting model in 5 minutes:  
[studio.edgeimpulse.com/evaluate](https://studio.edgeimpulse.com/evaluate)

FOMO: [edgeimpulse.com/fomo](https://edgeimpulse.com/fomo)

Documentation: [docs.edgeimpulse.com](https://docs.edgeimpulse.com)

# Thank you!

@aureleq