Make you IoT Smarter with Tensorflow Lite ...

... to Design the Future of Vertical Farming

@alexis0duque

Who am I?

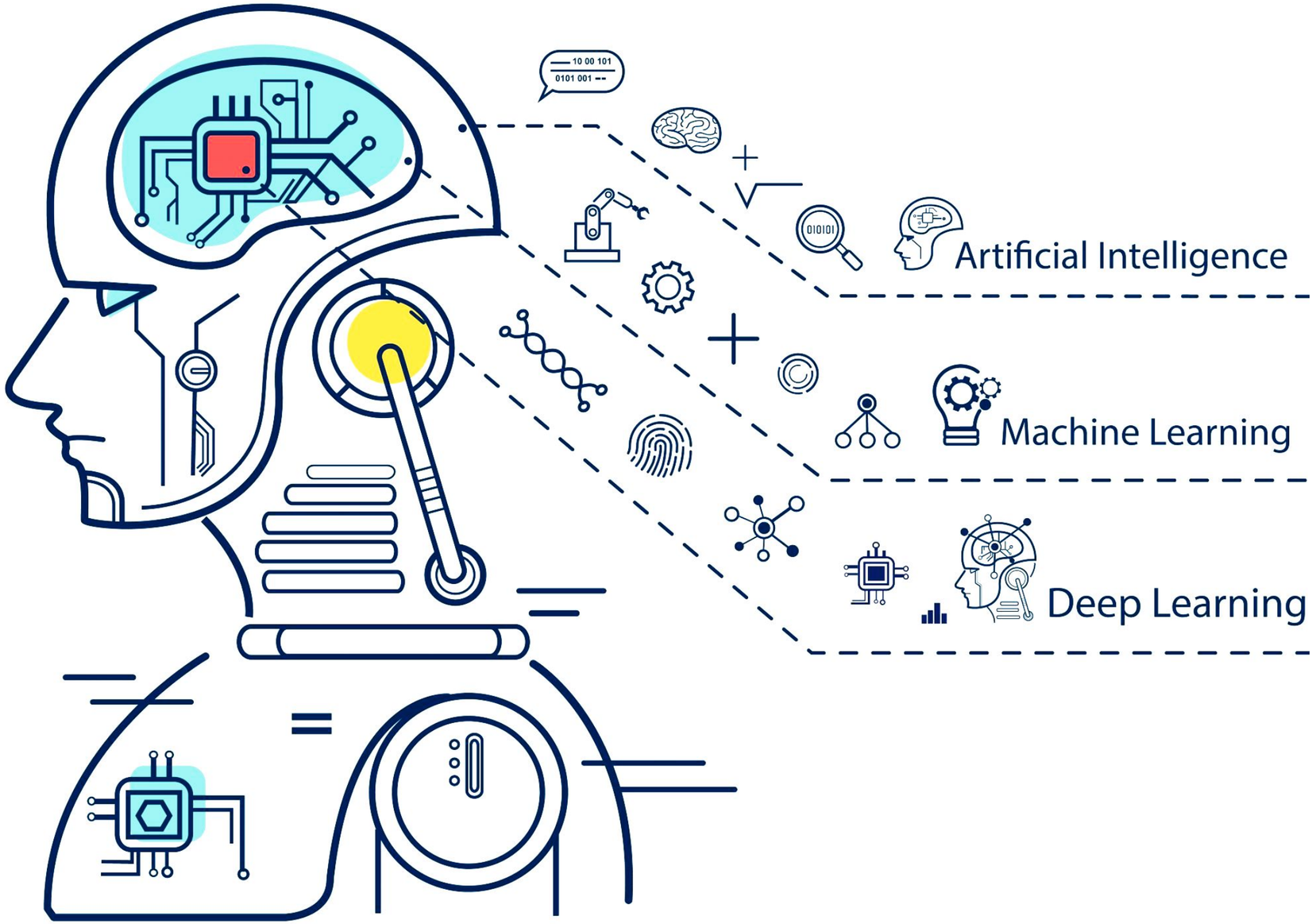
Alexis DUQUE

Director of Research & Development

- @alexis0duque
- alexisduque
- alexisd@rtone.fr
- alexisduque.me
- https://goo.gl/oNUWu6

RTONE
IOT MAKERS

DEVOXX TM


10 00 101
0101 001 --
Artificial Intelligence
Machine Learning
Deep Learning

?

DEVOXX™

TensorFlow Lite ?

DEVOXX™

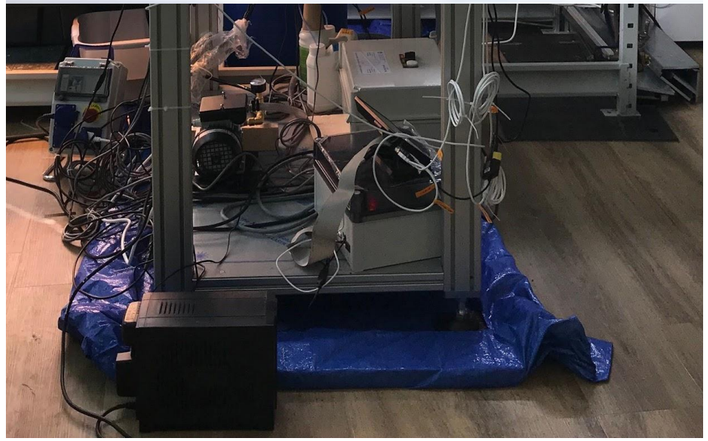
?

Outline & What You Will Learn

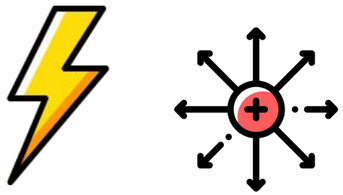
- **Indoor Vertical Farming**
- Why **Intelligence** at the **Edge**?
- Introduction to **Tensorflow Lite.** How to Use It?
- **Setup** your laptop & RPI
- **Build** and **train** a small model to **predict lettuce weight**
- **Convert** and deploy it on a RPI
- Run **predictions** on **IoT devices**
- Benchmarks
- Further Work

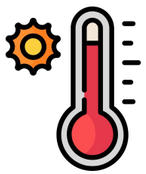
Indoor Vertical Farming

LA GRANGETTE



System Architecture





avg **co2** (double, in ppm)

avg **dissolved oxygen** (double, in ppm)

avg **electrical conductivity** (double, in µS/cm)

avg **RedOx potential** (double, in mV)

avg **PPFD** (Photosynthetic Photon Flux Density, double, in µmol/m2/s)

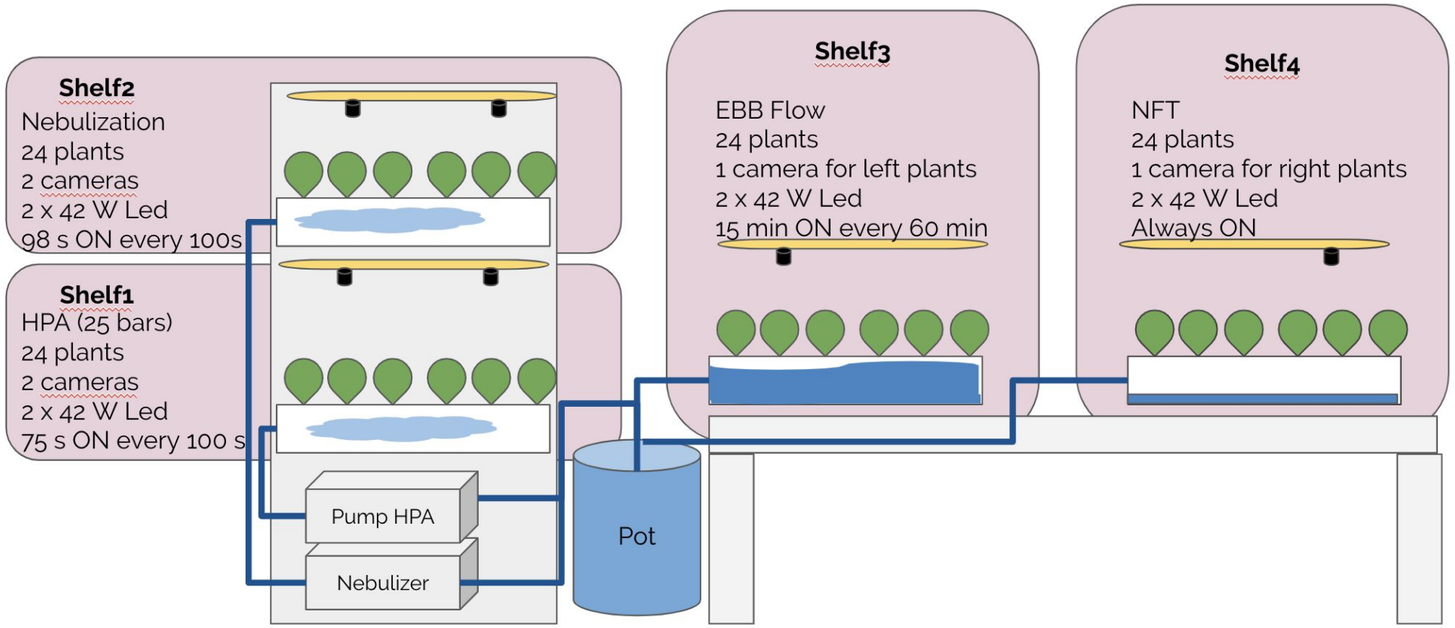
avg **water pH** (double, in -)

average **humidity** (double, in %)

average **temperature** (double, in °C)

System Architecture

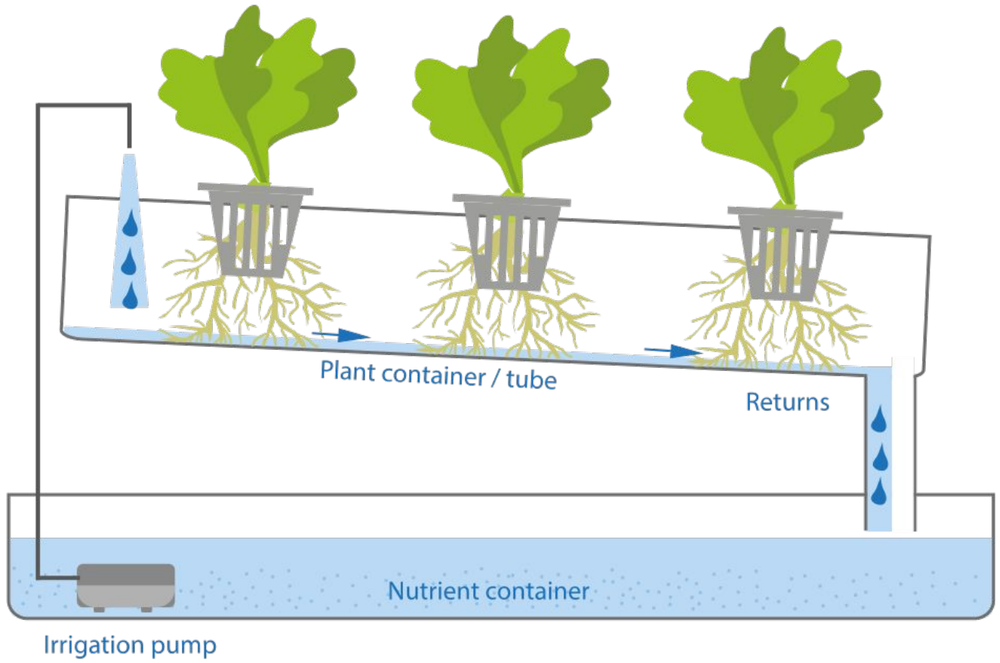
Actuators and Irrigation System

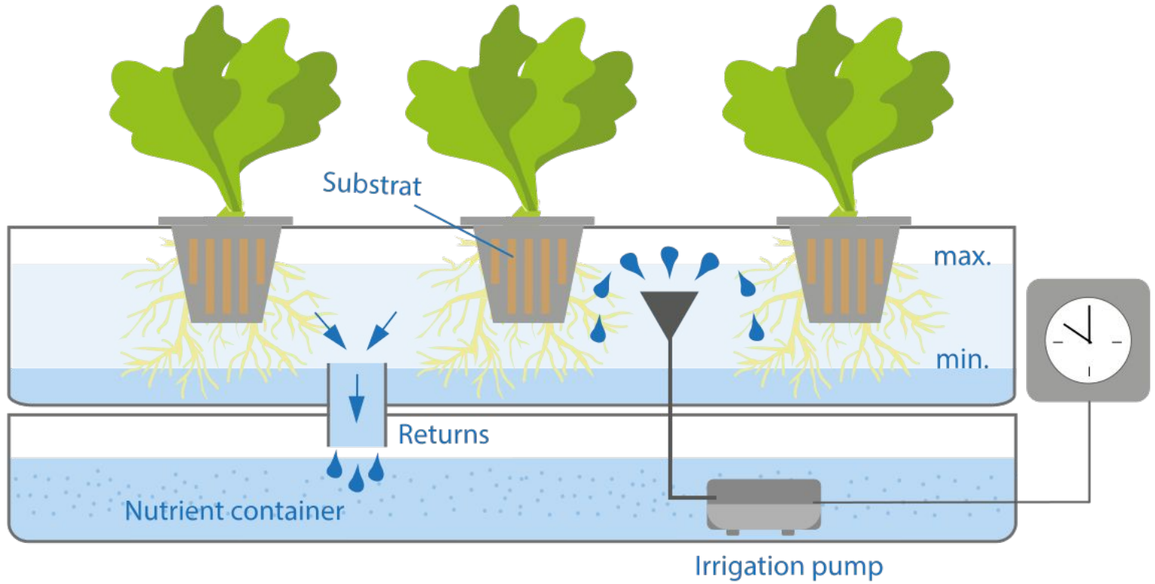


Sensors

Irrigation System

- HPA
- Nebulization
- NFT
- Ebb & Flow

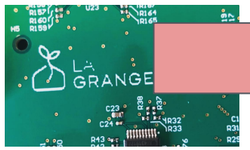




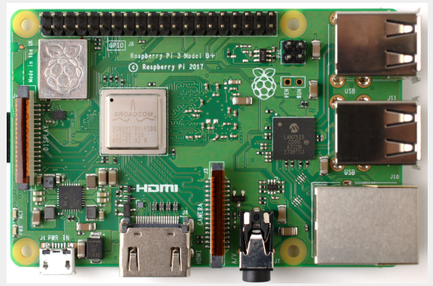
https://www.hydroponic-urban-gardening.com/hydroponics-guide/various-hydroponics-systems

System Architecture

Sensors



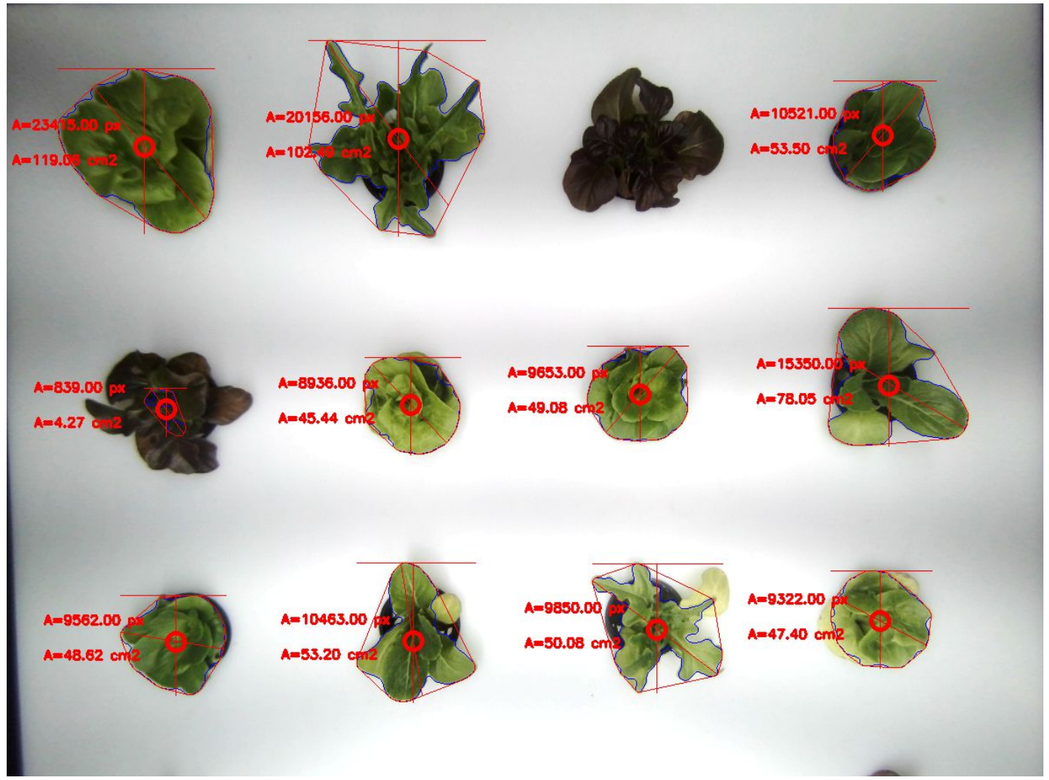
Motherboard

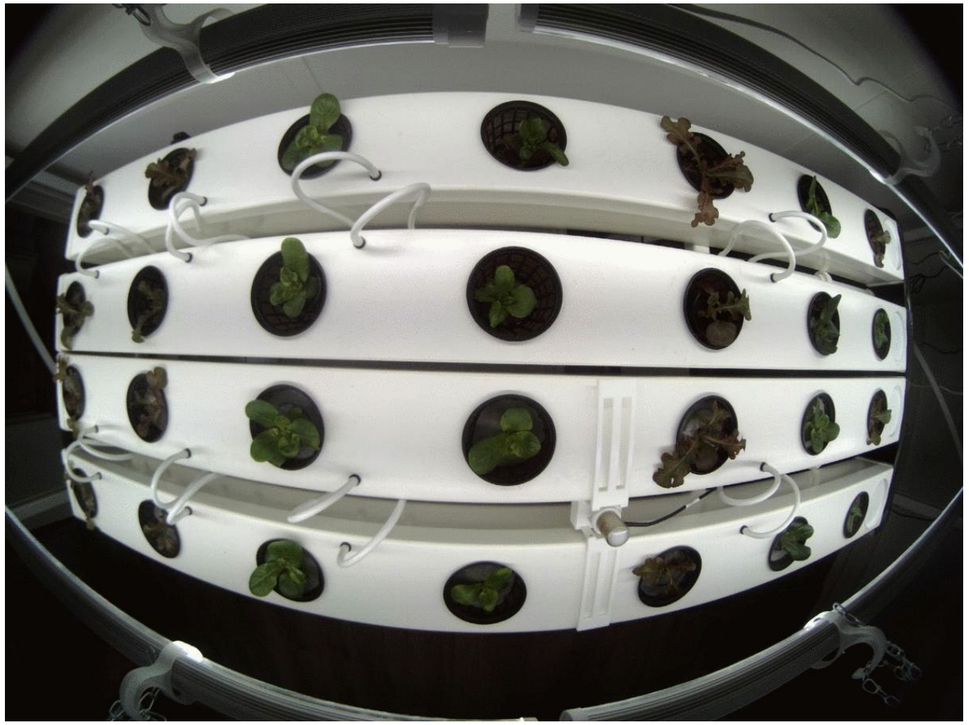


OpenCV

TensorFlow Lite

Computer Vision

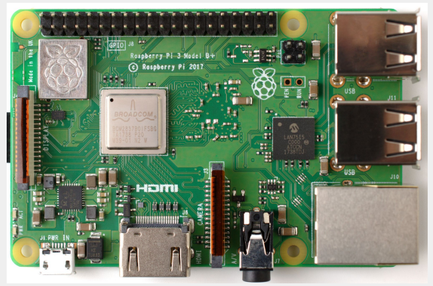




System Architecture

Cloud

Motherboard



OpenCV

TensorFlow Lite

LA GRANGETTE **Units** Plantings Actions Measurements Help

Units > 5

5 - Unit 5 - 2 shelves Opened - EBB Flow

Activity | Sensors | Edition

Unit sensors

- [ ] CO2 co2_atlas_02
- [ ] Dissolved Oxygen do_atlas_pot_02
- [ ] Electrical Conductivity ec_atlas_pot_02
- [ ] Oxydo-Reduction Potential orp_atlas_pot_02
- [x] PPFD par_sq500_01
- [ ] PPFD par_sq500_02
- [x] pH ph_atlas_02
- [ ] Relative Humidity hum_am2315_01
- [ ] Relative Humidity hum_am2315_02
- [x] Air Temp. temp_am2315_01
- [ ] Air Temp. temp_am2315_02
- [ ] Water Temp. temp_ds18b20_01

Action types

- irrigation
- light
- nutrition

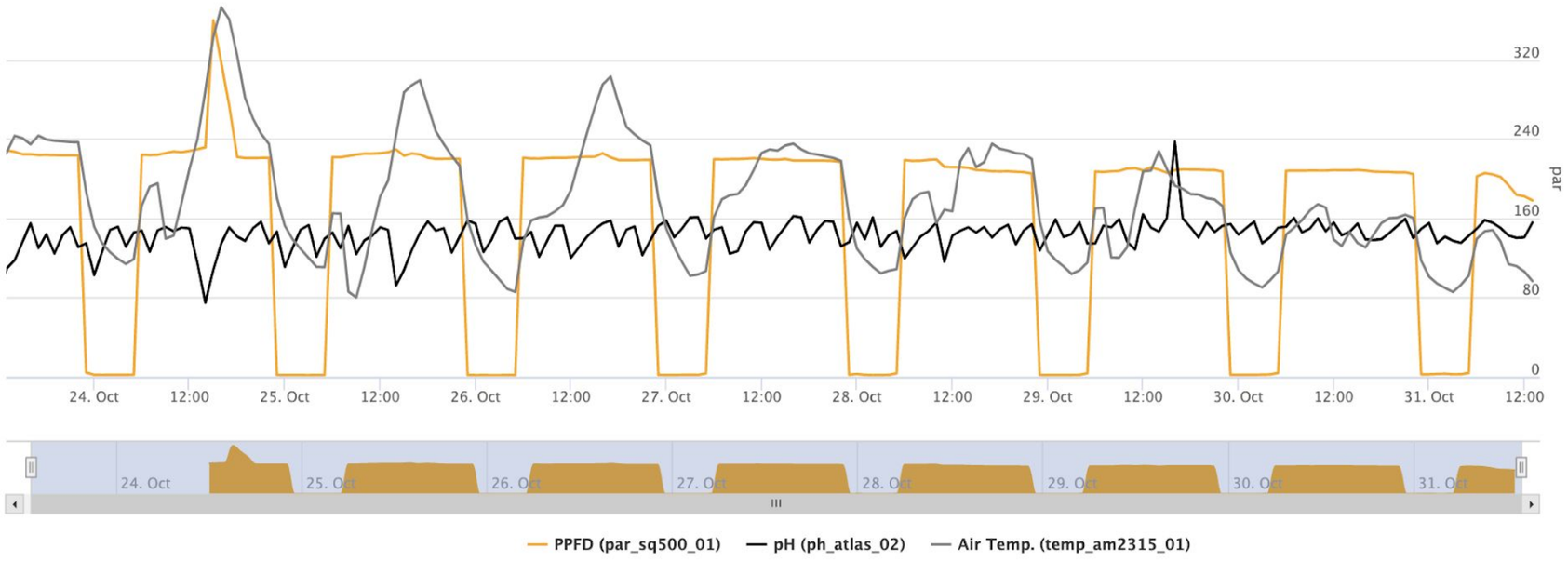
Start date 10/24/2019, 2:56 PM

End date 10/31/2019, 2:56 PM

Displayed: *average value per hour*

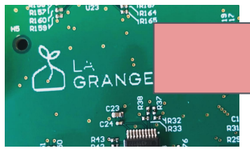
Zoom hour day week All

From Oct 23, 2019 To Oct 31, 2019



System Architecture

Sensors

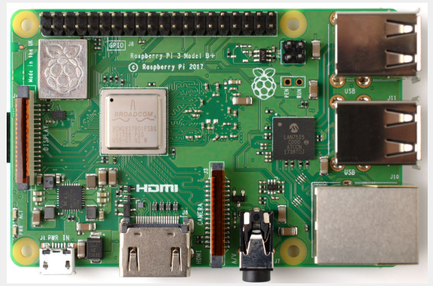


R&D

jupyter

Cloud

Motherboard



OpenCV

TensorFlow Lite

#DevoxxMA #IoT #Tensorflow #AI

@alexis0duque

Why Machine Learning?

Is the system working as expected?

When my lettuce will be ready to be eaten?

What should I do to make vegetables looks and tastes better?

What should I do to make it grow faster?

Flavor-cyber-agriculture: Optimization of plant metabolites in an open-source control environment through surrogate modeling

Johnson AJ, et al. (2019) Flavor-cyber-agriculture: Optimization of plant metabolites in an open-source control environment through surrogate modeling. PLOS ONE 14(4): e0213918.

Why Edge Computing?

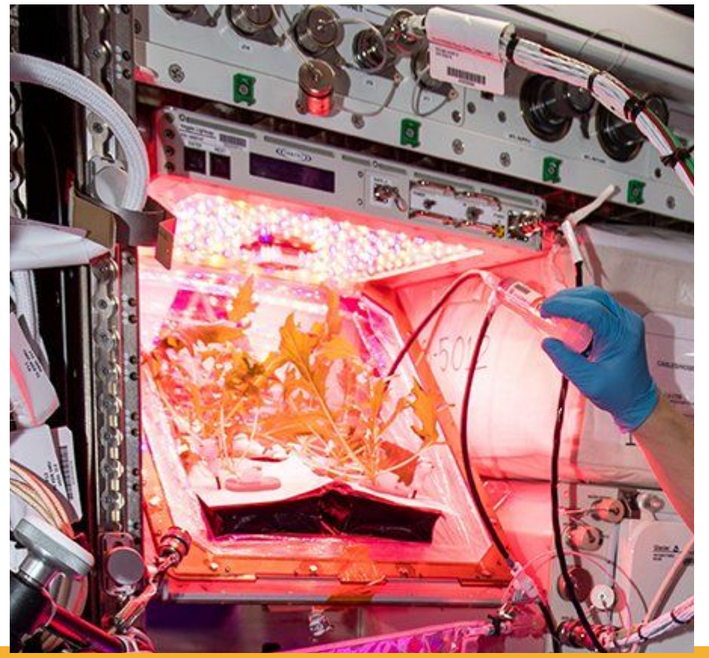
Infrastructure and cloud cost

Scalability

Must work in the field, without internet

Network Latency

Privacy



Tensorflow

TensorFlow

Open source library created by Google

Platform for Machine Learning

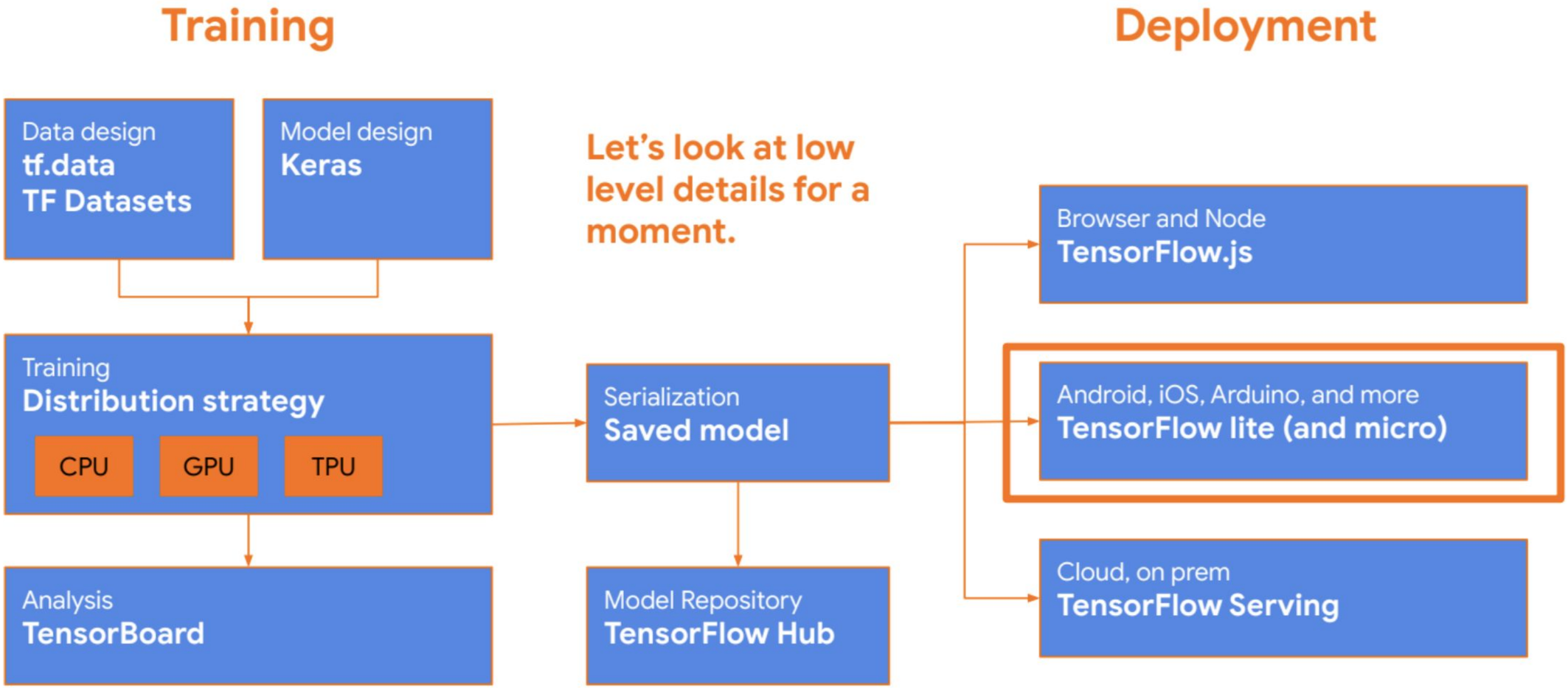
2.0 (released on October 2019)

Create, train, debug and use various machine learning model (neural network but not only!)

Keras, Lite, Tensorboard, Tensorflow Probability

Tensorflow

TensorFlow



Tensorflow Lite (TF-Lite)

TensorFlow Lite

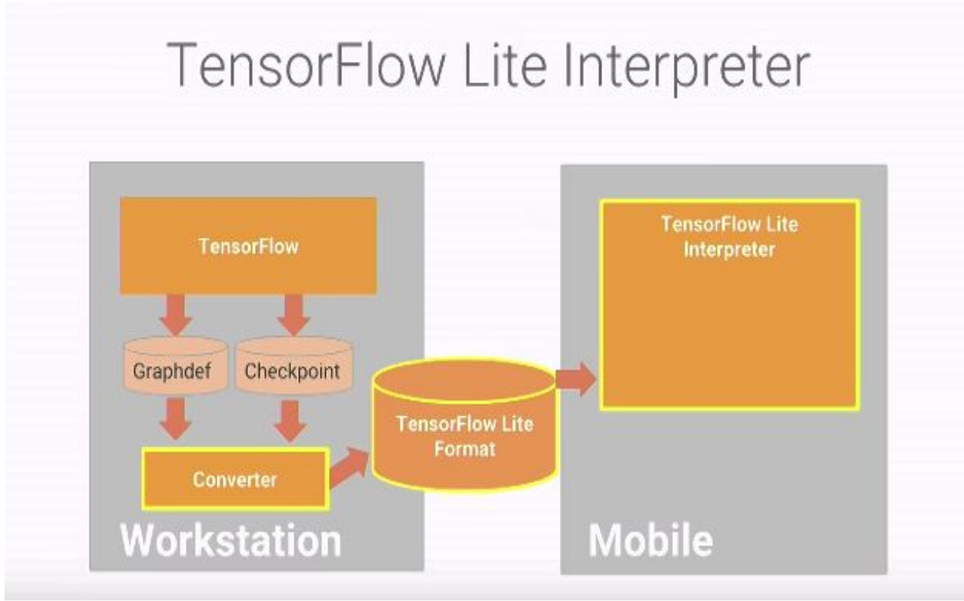
Converter + Interpreter

Tensorflow vs TF-Lite ?

- smaller model size
- faster inference
- mobile, embedded, MCU

but

- no training
- model is frozen => no re-training
- no transfer learning

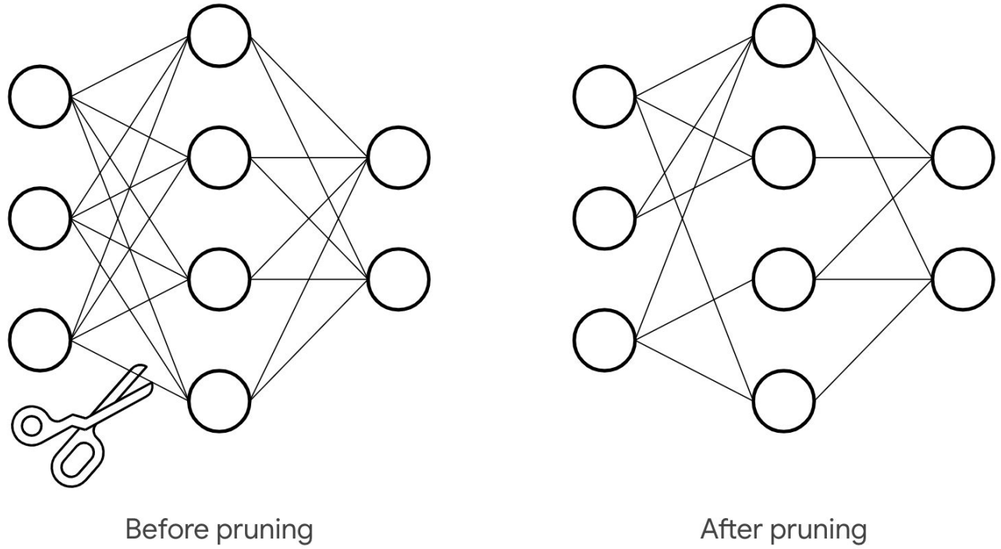


Tensorflow Lite (TF-Lite)

TensorFlow Lite

Optimization

- Pruning
- Post Training Quantization



Before pruning

After pruning

Delegate to offload execution

- GPU, TPU, DSP

ML Workflow with Tensorflow Lite

Import your **dataset**

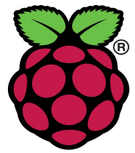
Work on data: preprocessing, normalization, features selection

Build your model with **Tensorflow**

Train your model

Export and **convert** to **.tflite**

ML Workflow with Tensorflow Lite




Load your model (*or grab one in* github.com/tensorflow/models*)*

Preprocess input data

Allocate Memory

Run inference

Interpret output

Setup on your Laptop

RtoneloT@devoxx

```
  Using cached https://files.pythonhosted.org/packages/0e/3e/377007e3f36ec42f1b84ec322ee12141a9e10d808312e5738f52f80a232c/pexpect-4.7.0-py2.py3-none-any.whl
Processing ./Library/Caches/pip/wheels/98/b0/dd/29e28ff615af3dda4c67cab719dd51357597eabff926976b45/backcall-0.1.0-cp37-none-any.whl
Collecting pickleshare
  Using cached https://files.pythonhosted.org/packages/9a/41/220f49aaea88bc6fa6cba8d05ecf24676326156c23b991e80b3f2fc24c77/pickleshare-0.7.5-py2.py3-none-any.whl
Collecting jedi>=0.10
  Using cached https://files.pythonhosted.org/packages/55/54/da994f359e4e7da4776a200e76dbc85ba5fc319eefc22e33d55296d95a1d/jedi-0.15.1-py2.py3-none-any.whl
Collecting wcwidth
  Using cached https://files.pythonhosted.org/packages/7e/9f/526a6947247599b084ee5232e4f9190a38f398d7300d866af3ab571a5bfe/wcwidth-0.1.7-py2.py3-none-any.whl
Collecting webencodings
  Using cached https://files.pythonhosted.org/packages/f4/24/2a3e3df732393fed8b3ebf2ec078f05546de641fe1b667ee316ec1dcf3b7/webencodings-0.5.1-py2.py3-none-any.whl
Collecting attrs>=17.4.0
  Using cached https://files.pythonhosted.org/packages/a2/db/4313ab3be961f7a763066401fb77f7748373b6094076ae2bda2806988af6/attrs-19.3.0-py2.py3-none-any.whl
Collecting importlib-metadata
  Using cached https://files.pythonhosted.org/packages/f6/d2/40b3fa882147719744e6aa50ac39cf7a22a913cbcba86a0371176c425a3b/importlib_metadata-0.23-py2.py3-none-any.whl
Processing ./Library/Caches/pip/wheels/bb/46/00/6d471ef0b813e3621f0abe6cb723c20d529d39a061de3f7c51/pyrsistent-0.15.4-cp37-cp37m-macosx_10_14_x86_64.whl
```

Setup on your Laptop

```
$ python3 --version
$ pip3 --version
$ virtualenv --version

$ virtualenv --system-site-packages -p python3 ./venv
$ source ./venv/bin/activate
$ pip install --upgrade pip
$ pip install --upgrade tensorflow=2.0
$ pip install numpy pandas jupyter jupyterlab notebook
matplotlib
```

Setup on your RPI

Tensorflow Lite Interpreter

1. Using pip and official TF release (not alway up to date)
2. Cross compile Tensorflow for ARMv7 on your laptop
3. Build Bazel and Tensorflow on your RPI (> 24h)
4. **Using pip and a *community built* .whl package**

https://github.com/PINTO0309/Tensorflowlite-bin

Setup on your RPI

RtonelaT@devoxx

```
Connecting to raw.githubusercontent.com (raw.githubusercontent.com)|151.101.120.133|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 1493151 (1.4M) [application/octet-stream]
Saving to: ‘tflite_runtime-2.0.0-cp37-cp37m-linux_armv7l.whl’

tflite_runtime-2.0.0-cp37-cp37m-linux_armv7l.whl                    0%[
                                                                                                                    ]       0  --
tflite_runtime-2.0.0-cp37-cp37m-linux_armv7l.whl                  100%[=================================================
==================================================================================================================>]   1.42M  8.
65MB/s    in 0.2s

2019-10-26 16:57:00 (8.65 MB/s) - ‘tflite_runtime-2.0.0-cp37-cp37m-linux_armv7l.whl’ saved [1493151/1493151]

(pi-devoxx-tflite) pi@raspberrypi:~ $ pip install --upgrade tflite_runtime-2.0.0-cp37-cp37m-linux_armv7l.whl
Looking in indexes: https://pypi.org/simple, https://www.piwheels.org/simple
Processing ./tflite_runtime-2.0.0-cp37-cp37m-linux_armv7l.whl
Requirement already satisfied, skipping upgrade: numpy>=1.12.1 in ./pi-devoxx-tflite/lib/python3.7/site-packages (from tflite
-runtime==2.0.0) (1.17.3)
Installing collected packages: tflite-runtime
Successfully installed tflite-runtime-2.0.0
(pi-devoxx-tflite) pi@raspberrypi:~ $ python
Python 3.7.3 (default, Apr  3 2019, 05:39:12)
[GCC 8.2.0] on linux
Type "help", "copyright", "credits" or "license" for more information.
>>> import tflite_runtime
```

Setup on your RPI

Tensorflow Lite Interpreter

```
$ sudo apt install swig libjpeg-dev zlib1g-dev python3-dev
python3-numpy unzip
$ wget
https://github.com/PINTO0309/TensorflowLite-bin/raw/master/2
.0.0/tflite_runtime-2.0.0-cp37-cp37m-linux_armv7l.whl
$ pip install --upgrade
tflite_runtime-2.0.0-cp37-cp37m-linux_armv7l.whl
```

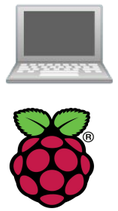
Setup on your RPI

Full Tensorflow package

```
sudo apt-get install -y libhdf5-dev libc-ares-dev libeigen3-dev
pip install keras_applications==1.0.8 --no-deps
pip install keras_preprocessing==1.1.0 --no-deps
pip install h5py==2.9.0
sudo apt-get install -y openmpi-bin libopenmpi-dev
sudo apt-get install -y libatlas-base-dev
pip install six wheel mock
wget
https://github.com/PINTO0309/Tensorflow-bin/raw/master/tensorflow-2.0.0-cp37-cp37m-linux_armv7l.whl
pip uninstall tensorflow
pip install tensorflow-2.0.0-cp37-cp37m-linux_armv7l.whl
```

DEVOXX™

Demo

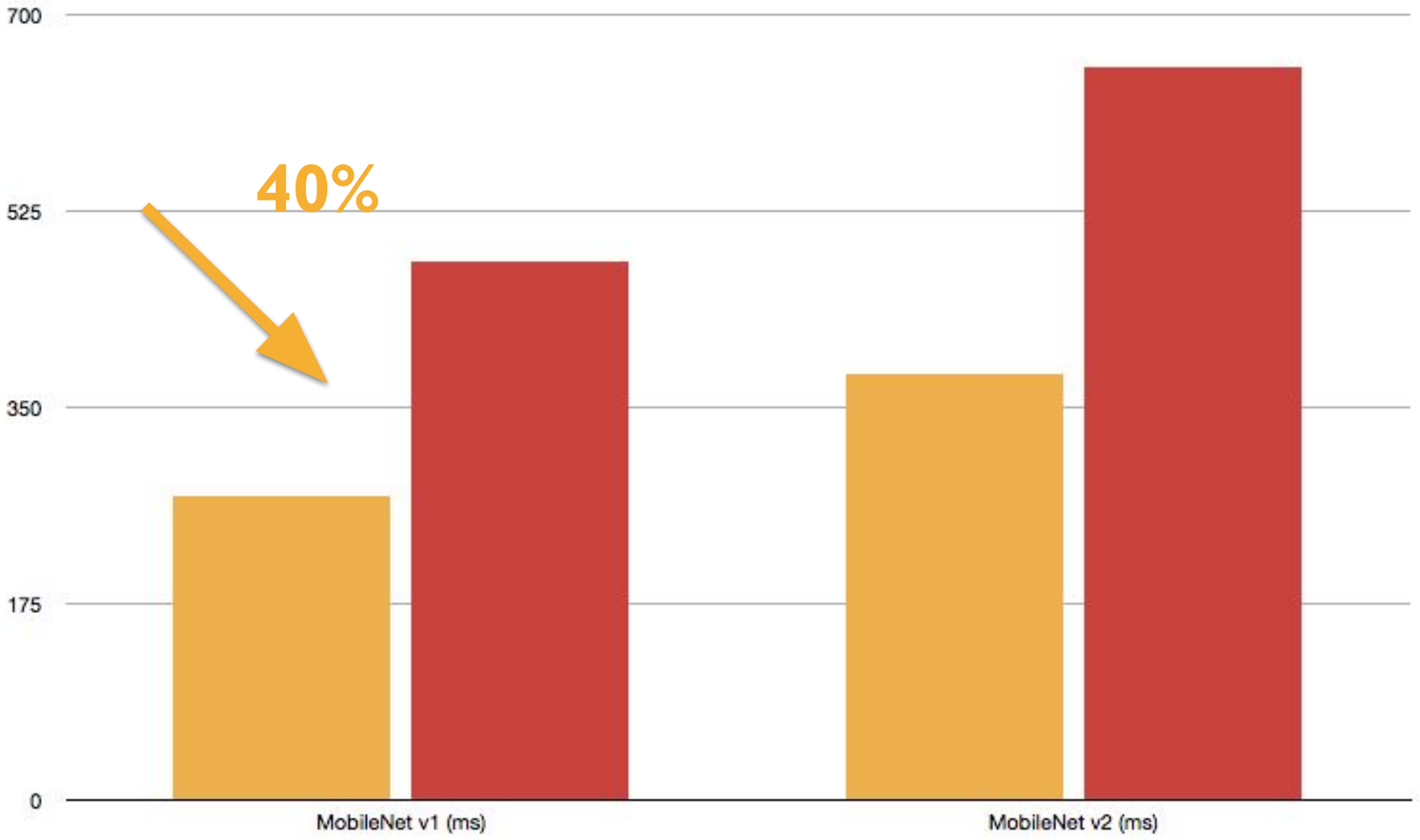


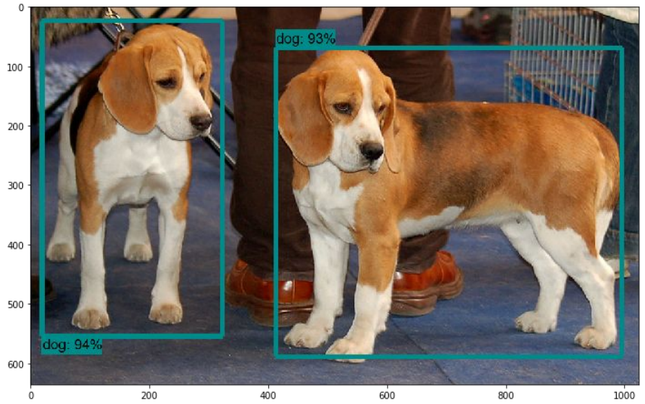
Part 1 - Build, Train and Convert a simple Neural Network model to predict lettuce weight

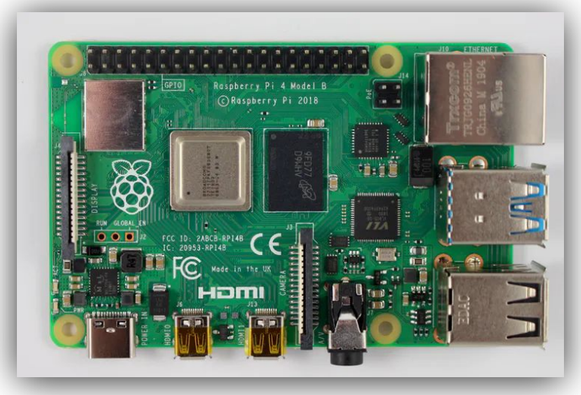
Part 2 - Deploy your tflite model on RPI and run inference

Tensorflow Lite Benchmark

Inference Time



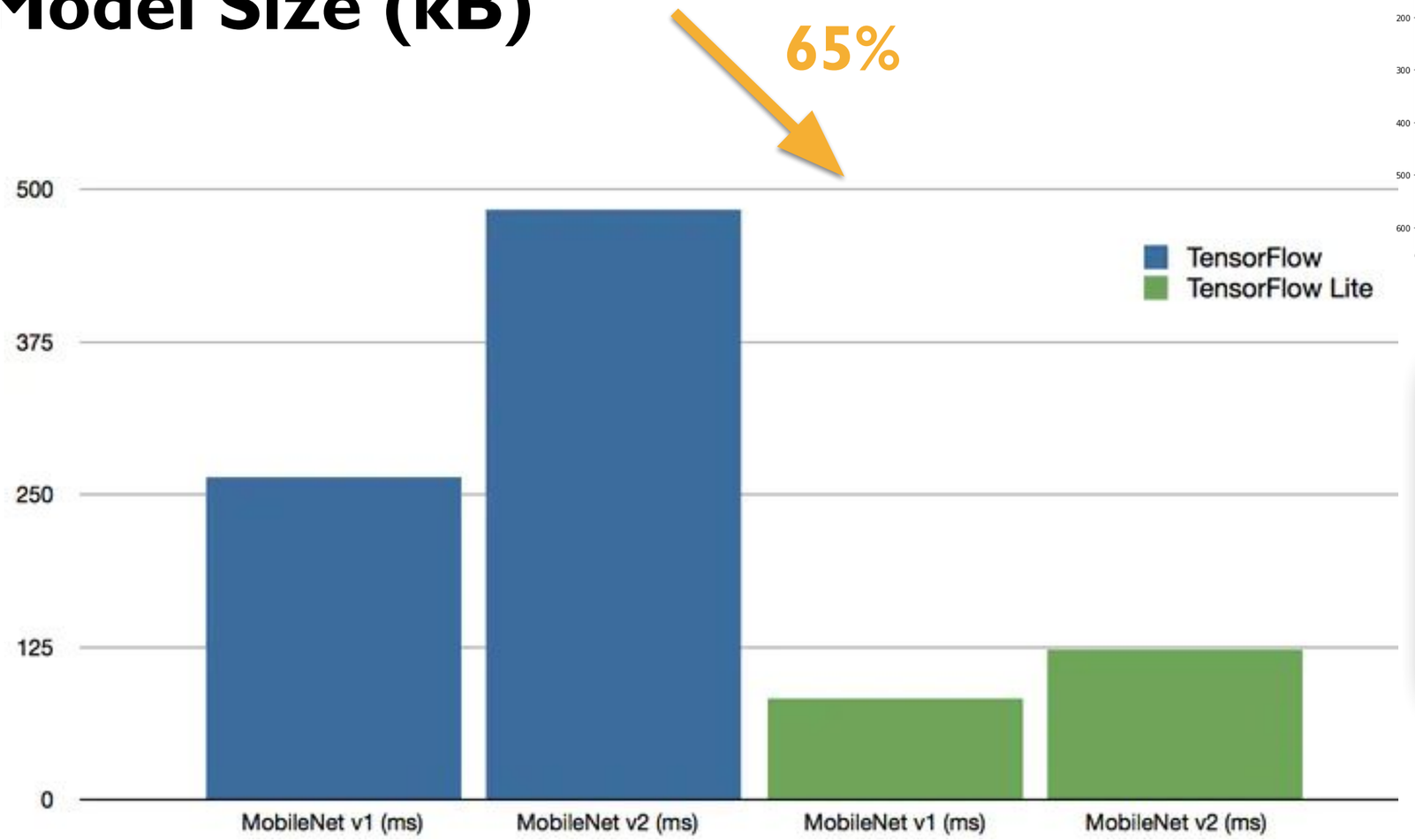


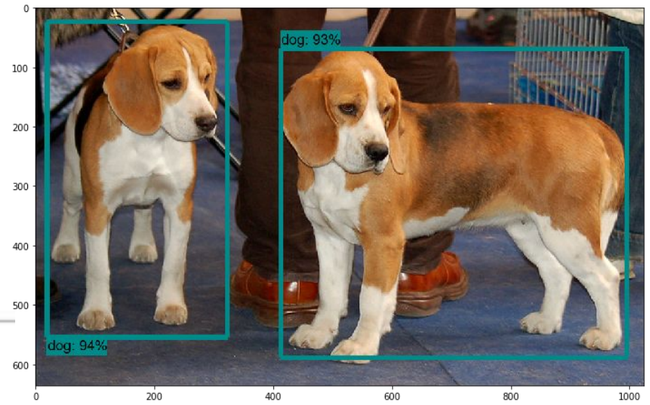


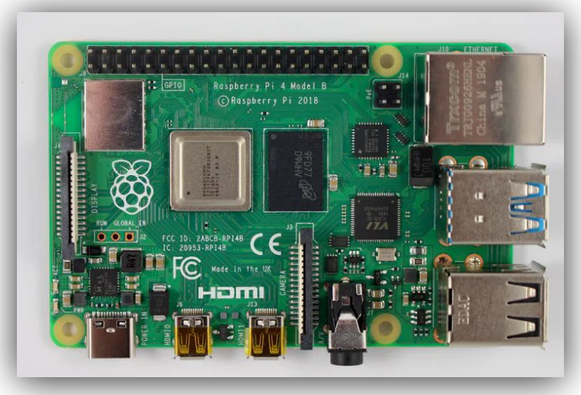
Source: Alasdair Allan

Tensorflow Lite Benchmark

Model Size (kB)







Source: Alasdair Allan

Tensorflow Lite Limitations

Reinforcement Learning

Transfer Learning

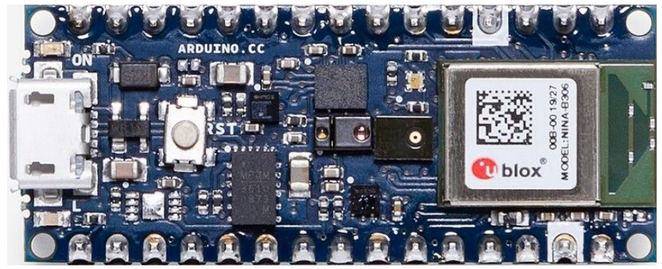
Recurrent Neural Network (RNN) like LSTM

Operation Compatibility

https://www.tensorflow.org/lite/guide/ops_compatibility

https://github.com/tensorflow/tensorflow/blob/master/tensorflow/lite/experimental/examples/lstm/g3doc/README.md

TFLite on Microcontrollers



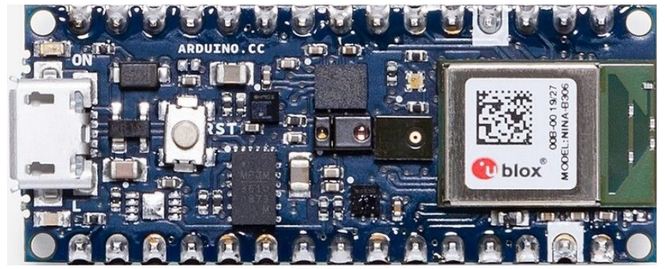
Inference on Cortex-M microcontroller

Arduino Nano 33

Only some operations are supported

Enough for hotword, gesture and speech recognition

TFLite on Microcontrollers

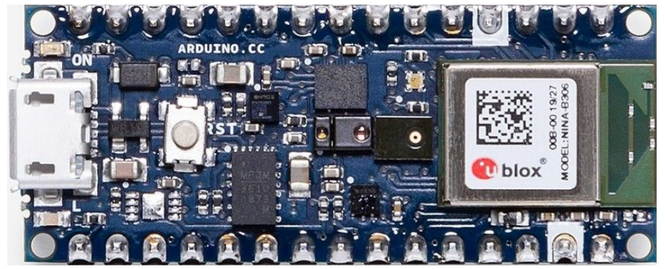


C++ API

```
#include <TensorFlowLite.h>
// This is your tflite model
#include "lg_weight_model.h"
#include
"tensorflow/lite/experimental/micro/kernels/all_ops_resolver.h"
#include "tensorflow/lite/experimental/micro/micro_interpreter.h"
#include "tensorflow/lite/schema/schema_generated.h"

tflite::ErrorReporter *error_reporter = nullptr;
const tflite::Model *model = nullptr;
tflite::MicroInterpreter *interpreter = nullptr;
TfLiteTensor *input = nullptr;
TfLiteTensor *output = nullptr;
```

TFLite on Microcontrollers

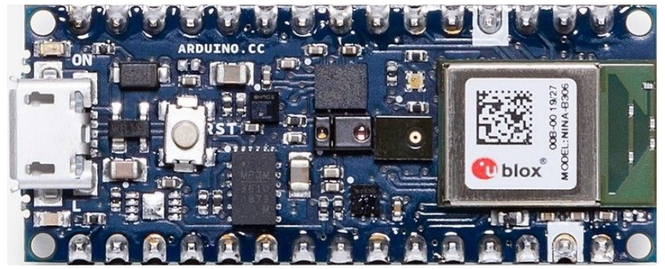



```
// Finding the min value for your model may require tests!
constexpr int kTensorArenaSize = 2 * 1024;
uint8_t tensor_arena[kTensorArenaSize];

// Load your model.
model = tflite::GetModel(g_weight_regresion_model_data);
// This pulls in all the operation implementations we need.
static tflite::ops::micro::AllOpsResolver resolver;

// Build an interpreter to run the model with.
static tflite::MicroInterpreter static_interpreter(
  model, resolver, tensor_arena, kTensorArenaSize, error_reporter);
interpreter = &static_interpreter;
```

TFLite on Microcontrollers




```
// Allocate memory for the model's tensors.
TfLiteStatus allocate_status = interpreter->AllocateTensors();
// Obtain pointers to the model's input and output tensors.
input = interpreter->input(0);
output = interpreter->output(0);
// Feed the interpreter with the input value
float x_val = random(0, 10);
input->data.f[0] = x_val;

// Run Inference
TfLiteStatus invoke_status = interpreter->Invoke();
// Get inference result
float y_val = output->data.f[0];
```

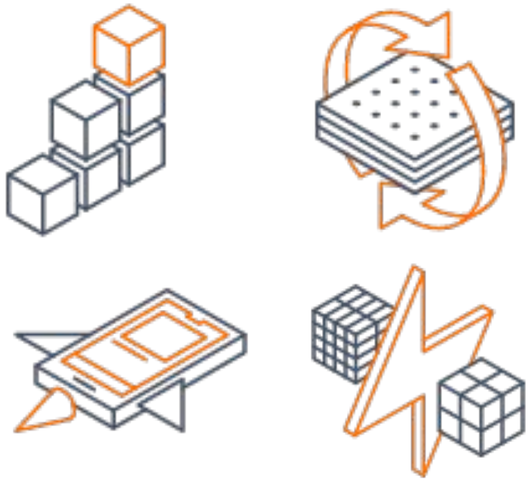
Further Work

Training at the Edge

Transfer Learning

Federated Learning

Summary



Build, Train, Optimize, Convert on laptop

Deploy, Infer on device

Some operations are not supported

Quantization does not affect accuracy

Inference on IoT and microcontrollers is feasible

Regression, anomalies detection, objects recognition, smart reply, etc.

DEVOXX™

Thanks!

frama.link/tflite-Devoxx

frama.link/rtone-jobs

References

https://medium.com/tensorflow/how-to-get-started-with-machine-learning-on-arduino-7daf95b4157

https://www.tensorflow.org/lite

https://www.tensorflow.org

https://coral.withgoogle.com/

https://arxiv.org/abs/1902.01046

https://medium.com/tensorflow/tensorflow-model-optimization-toolkit-pruning-api-42cac9157a6a