

Running AI models on the NVIDIA Jetsons CPU and GPU with balena



Alan Boris and Marc Pous

balena.io



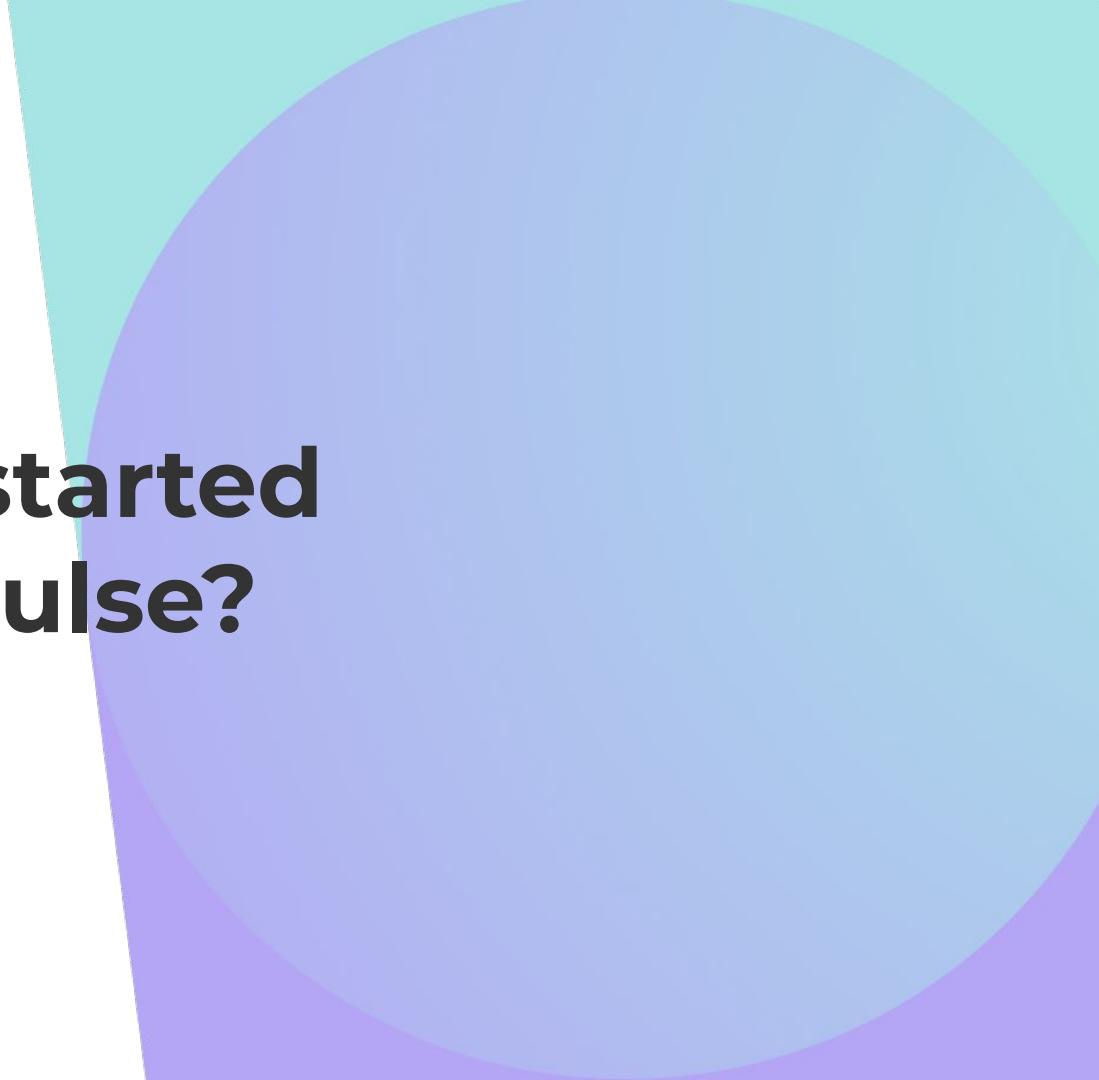
Alan Boris

Hardware Hacker at balena.io

alan@balena.io

<https://github.com/alanb128/>

@balena_io



How did I get started with Edge Impulse?



balena +  **EDGE IMPULSE**



Building an AI-Powered Bird Watcher with balena and Edge Impulse

Mithun Das

Edge Impulse Ambassador

Marc Pous

Developer Advocate at balena.io

01/10/2021

Copyright © Edge Impulse Inc.

EDGE IMPULSE
Imagine



<https://www.youtube.com/watch?v=md-X4o4q9il>

#2021Imagine

EDGE IMPULSE
Imagine



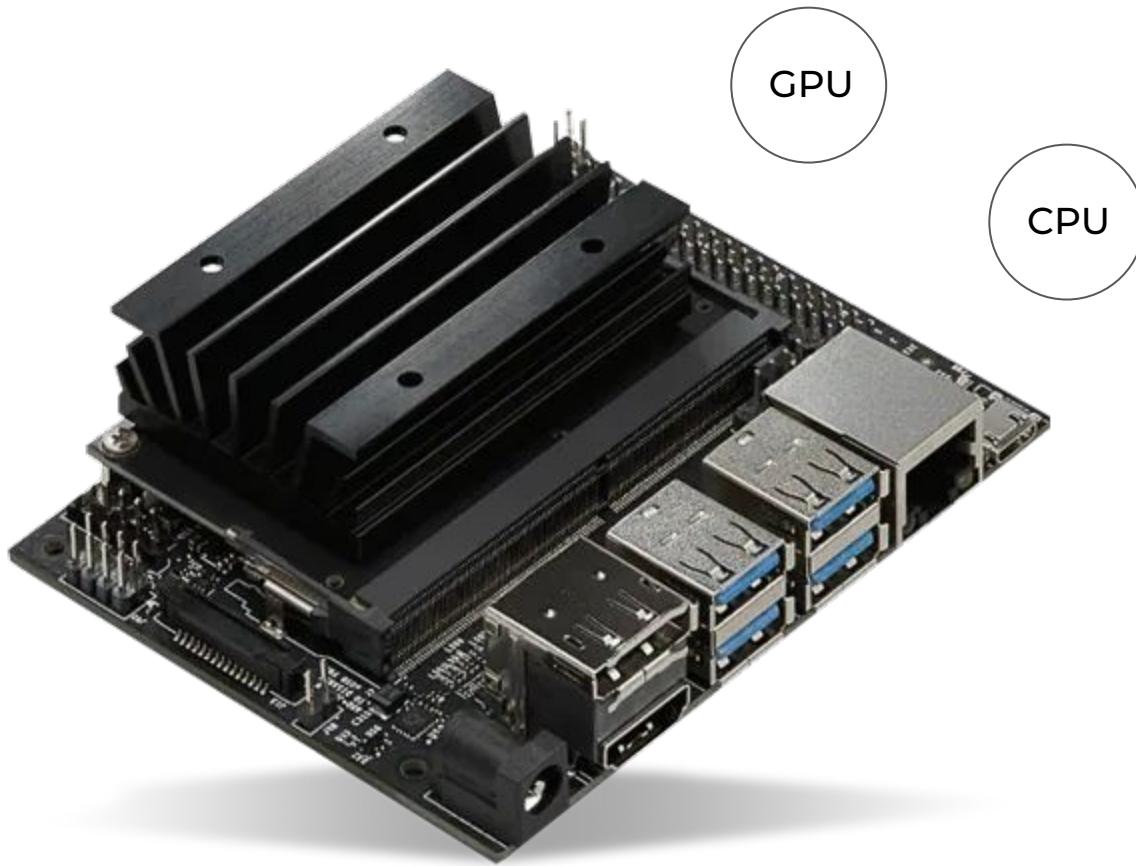
Alan

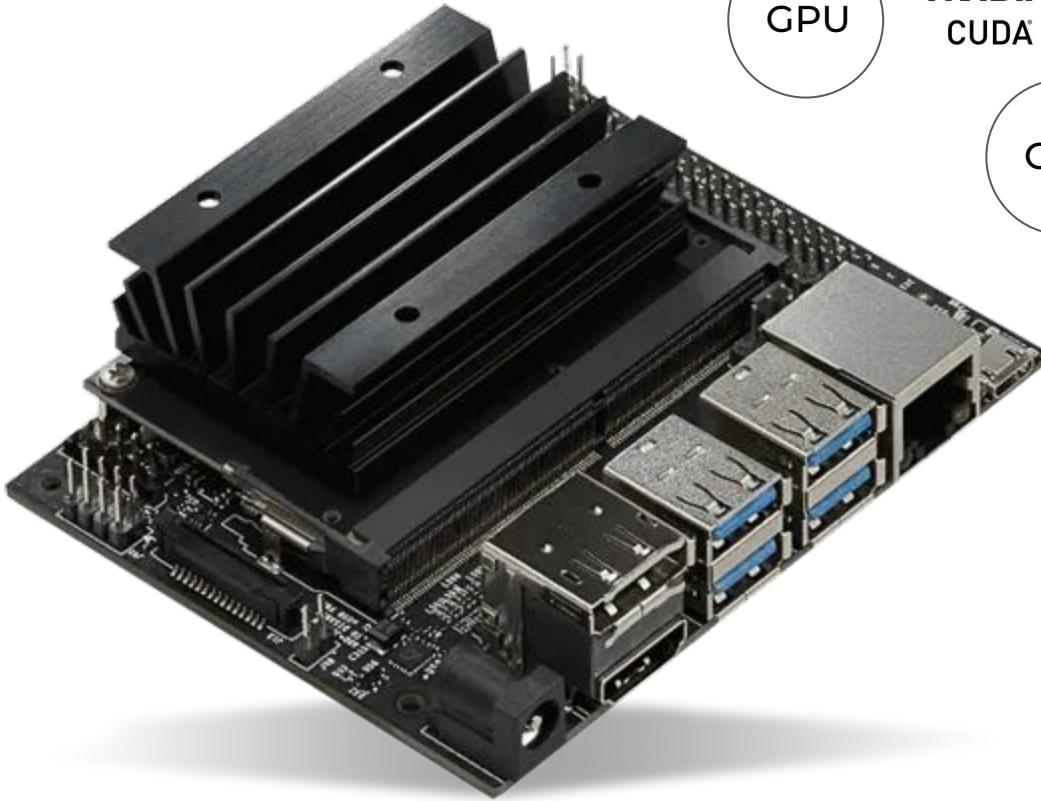
<https://youtu.be/Q0A343Q8Xlw?t=2694>







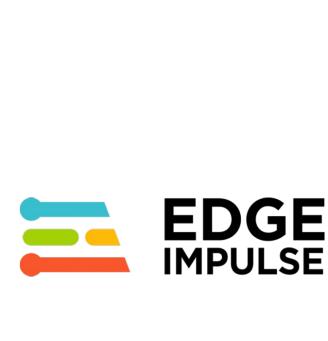




GPU



CPU

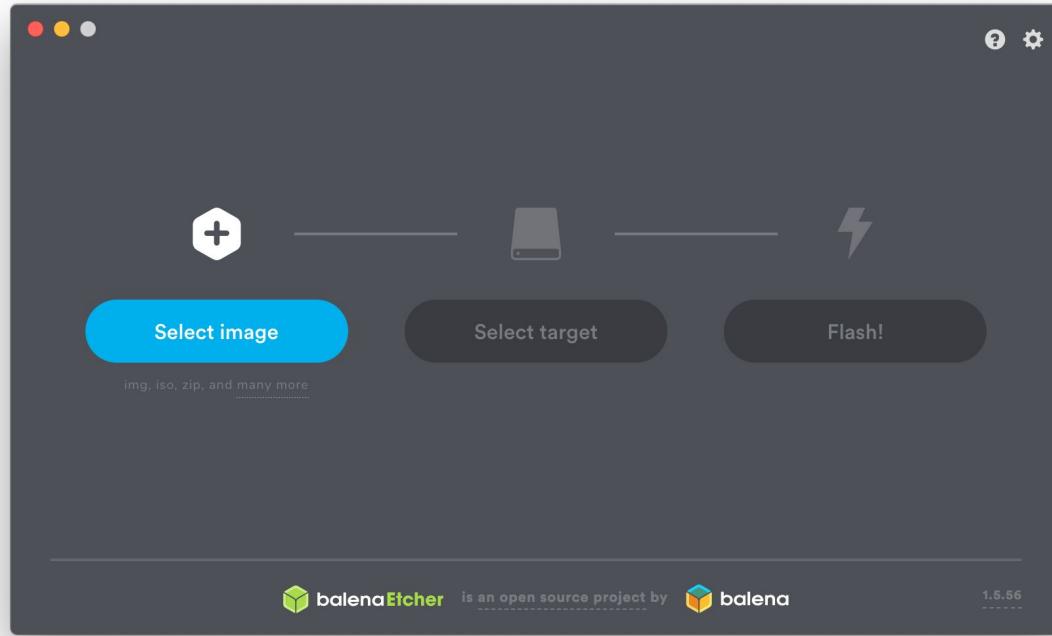


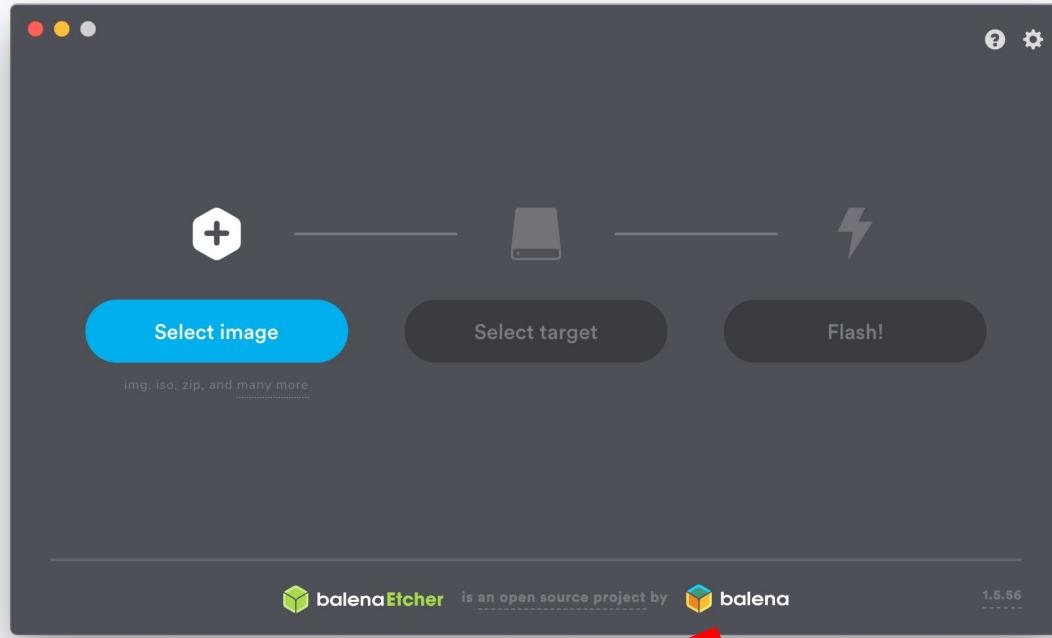
IMAGINE
EDGE IMPULSE



**And not only 1,
but multiple
deployed all
around the
world?**

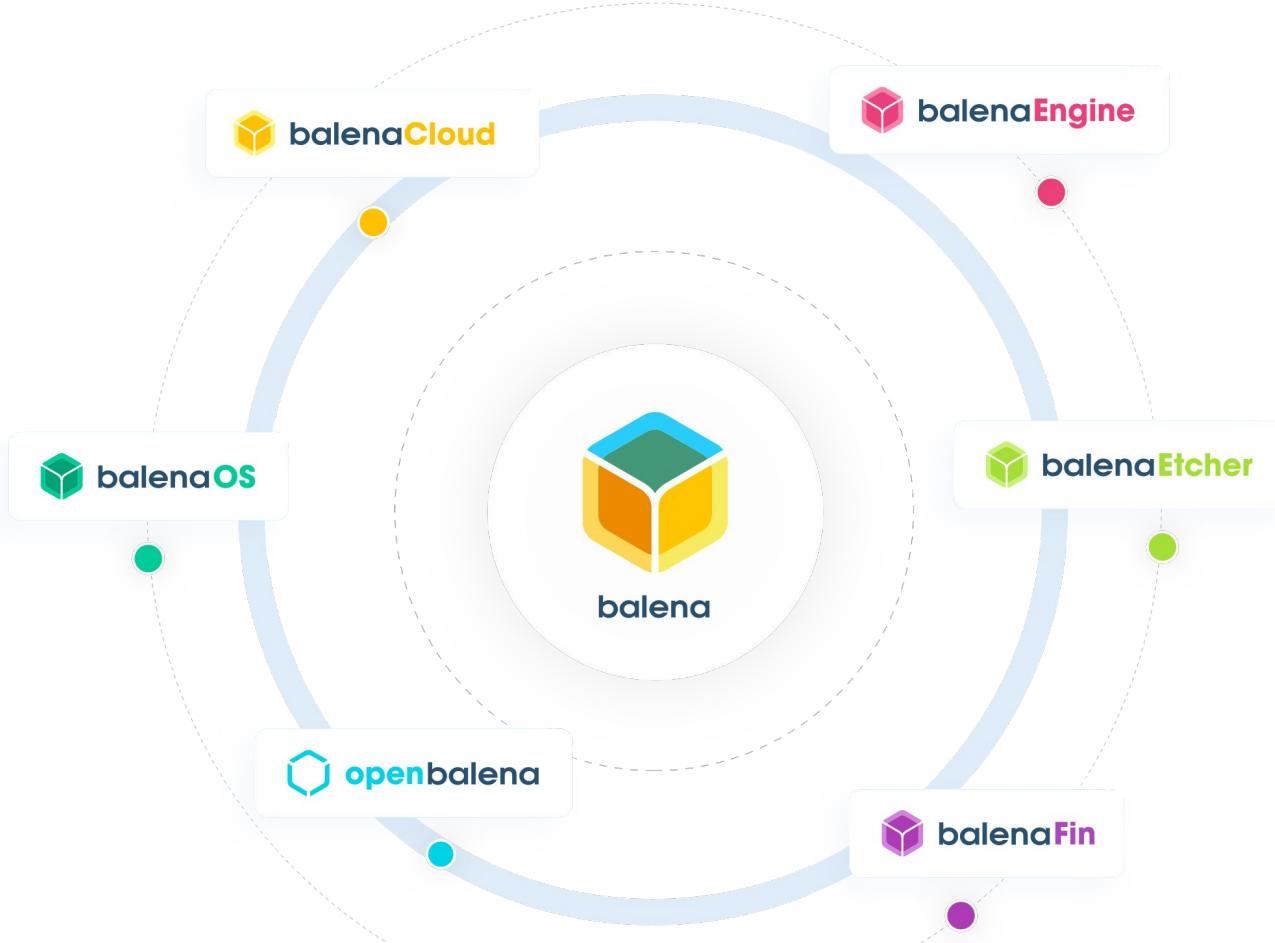
What is balena?

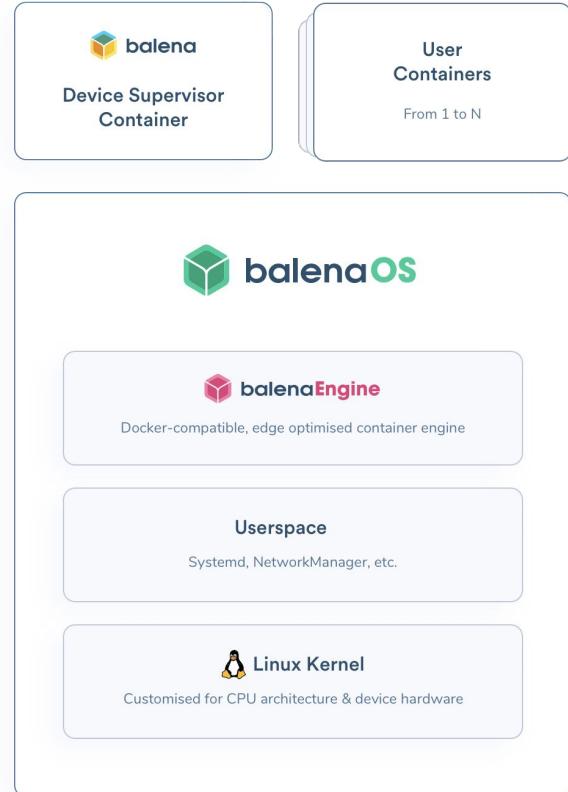




balena is:

- A platform for device fleet management.
- An operating system optimized for running Docker containers on embedded devices.
- Open source end-to-end.
- A cloud dashboard to keep track of all your devices.

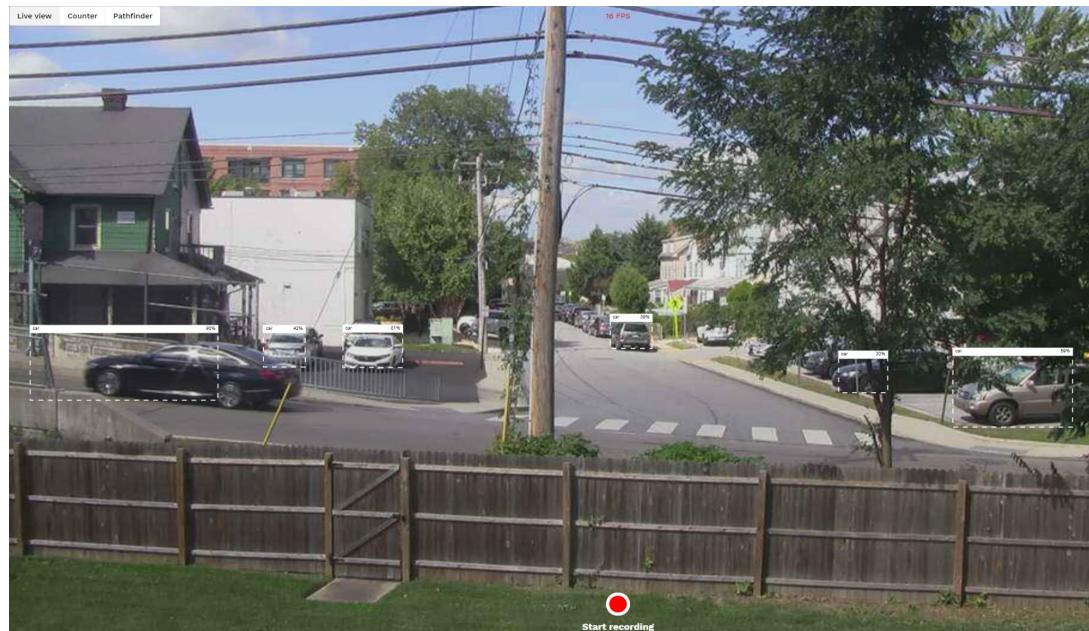




How we
Run AI models on the
NVIDIA Jetson's CPU and
GPU with balena

OpenDataCam: An open source tool to quantify the world

- Object detection algorithm that counts and tracks moving objects
- Runs on Jetson GPU using CUDA
- Outputs data, not images
- Runs fully on the Edge, no cloud connection
- Data available via a local API



balenalabs-incubator/opendatacam

https://github.com/balenalabs-incubator/opendatacam

Search or jump to... Pull requests Issues Marketplace Explore

Code Issues 2 Pull requests 1 Actions Projects Wiki Security Insights

master 5 branches 0 tags Go to file Add file Code

alanb128 Delete opencv411_lib.tar.gz a9e8d2c 8 days ago 106 commits

Dockerfile.jetson-nano Fix build and config issues 11 months ago

Dockerfile.jetson-nano-2gb-devkit Fix build and config issues 11 months ago

Dockerfile.jetson-tx2 Fix build and config issues 11 months ago

Dockerfile.jetson-xavier Fix build and config issues 11 months ago

Makefile.jetson-nano Update Makefile.jetson-nano 2 years ago

Makefile.jetson-tx2 Update Makefile.jetson-tx2 2 years ago

Makefile.jetson-xavier Update Makefile.jetson-xavier 2 years ago

README.md Update README.md 11 months ago

balena.yml Update balena.yml 2 years ago

config.jetson-nano Bump Opendatacam version to 3.0.2 12 months ago

config.jetson-tx2 Fix build and config issues 11 months ago

config.jetson-xavier Fix build and config issues 11 months ago

docker-compose.yml Update docker-compose.yml 8 days ago

launch.sh Update launch.sh 16 months ago

logo.png Add files via upload 2 years ago

opencv411_include.tar.gz Merge 3.0.1 branch (#7) 2 years ago

opencv411_lib.tar.gz Merge 3.0.1 branch (#7) 2 years ago

About

An implementation of OpenDataCam that is fully containerized and runs on balena + Jetson Nano.

Readme 25 stars 5 watching 16 forks

Releases

No releases published Create a new release

Packages

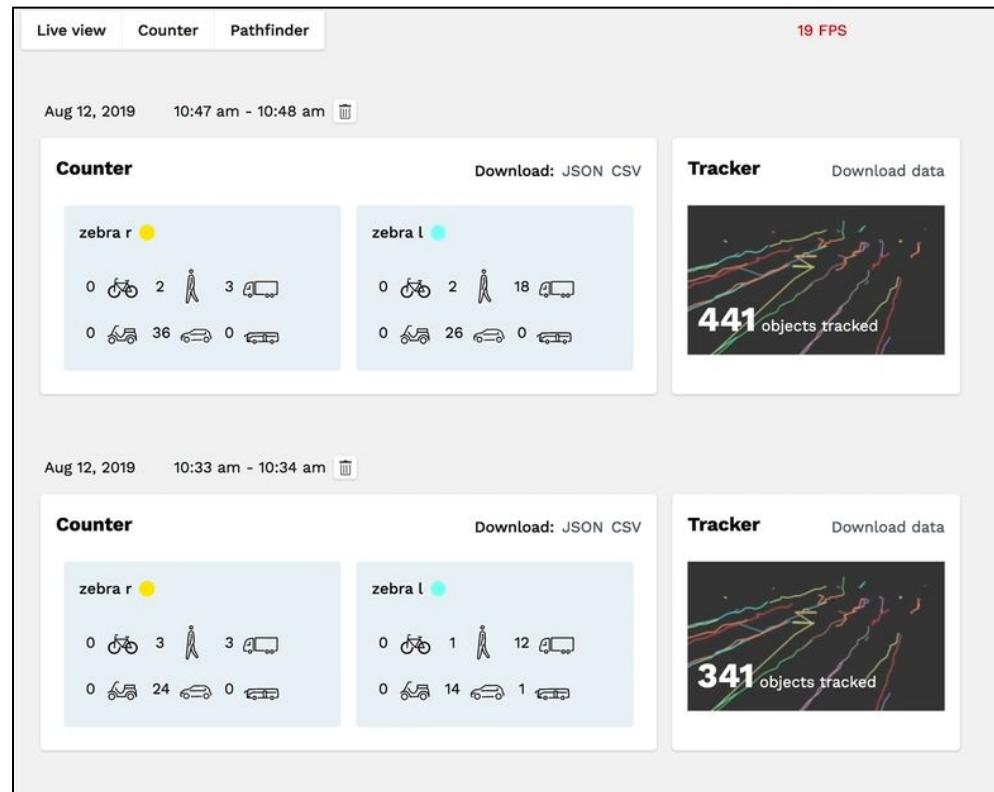
No packages published Publish your first package

Languages

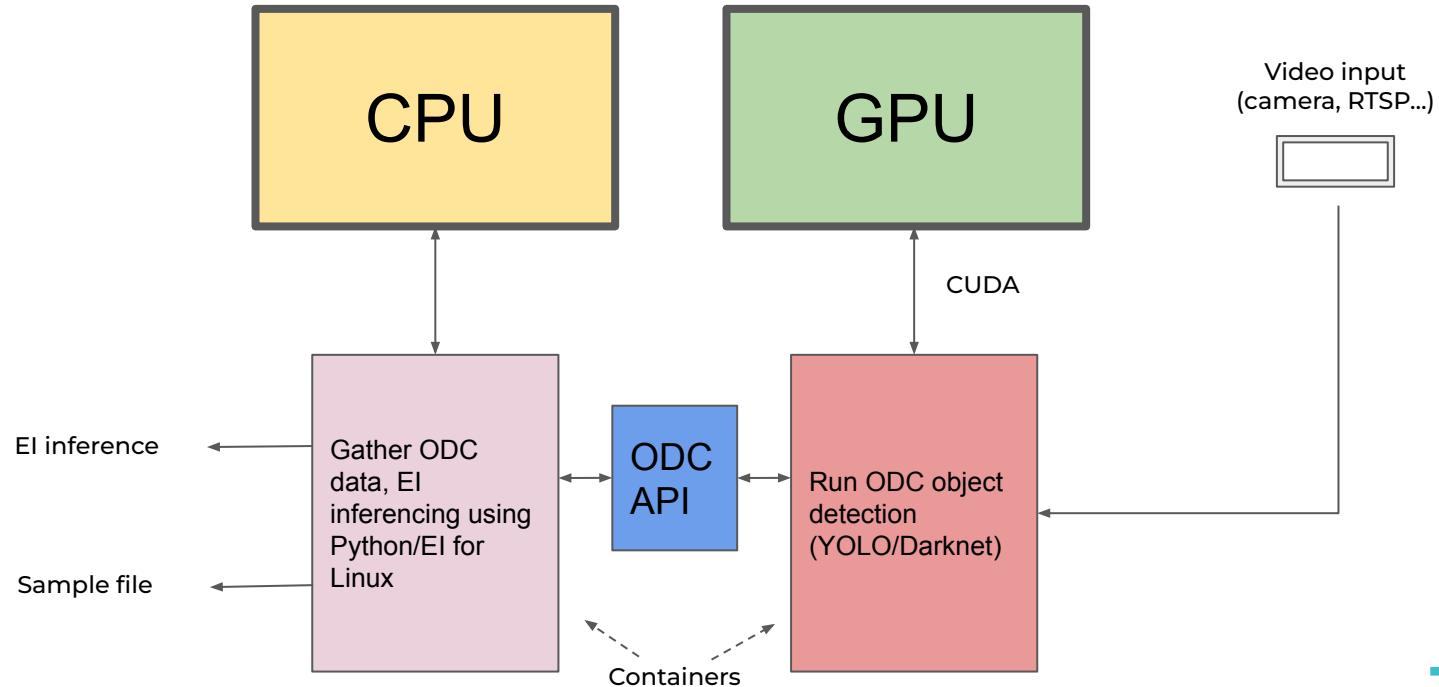
Shell 100.0%

OpenDataCam → Edge Impulse

- “Record” 60 second chunks of data on cars moving between two lines
- Using OpenDataCam API, feed this data to an EdgeImpulse model on the device
- Get an inference, then start over!
- Optionally, label the data chunk and send it back to EI cloud for re-training the model.



Different AI on Jetson Nano CPU and GPU





**Let's build it
together!**



alanb128/opendatacam-ei: Jetson Nano / Xavier Dev Board

https://github.com/alanb128/opendatacam-ei

Search or jump to... Pull requests Issues Marketplace Explore

Watch 1 Fork 0 Star 0

Code Issues Pull requests Actions Projects Wiki Security Insights

main 2 branches 0 tags Go to file Add file Code About

alanb128 Update runner.py 5c6ce18 2 hours ago 40 commits

ei Update runner.py 2 hours ago

opendatacam odc api pull 8 days ago

README.md Update README.md 2 days ago

balena.yml Upload Opendatacam last month

docker-compose.yml odc api pull 8 days ago

logo.png Upload Opendatacam last month

Readme 0 stars 1 watching 0 forks

About

Jetson Opendatacam on gpu and Edge Impulse on CPU

Releases

No releases published Create a new release

Packages

No packages published Publish your first package

Languages

Python 81.5% Shell 18.5%

README.md

OpenDataCam

An implementation of [OpenDataCam 3.0.1](#) that is fully containerized and runs on balena + Jetson Nano/TX2 /Xavier.

For Edge Impulse Imagine 2022 Demo

This version runs EI on the CPU and Opendatacam on the GPU

For this demo, you must start the Opendatacam UI first by browsing to local IP:8080

The script must manually be started in the ei service by running `runner.py` in `/usr/src/app`

The EI service starts data recordings of opendatacam and then runs inferences on them. The model used is determined by the API_KEY and HMAC_KEY, which causes the specified model to be downloaded at the container start. Optionally, the resulting data files can be uploaded back up to EI for re-training the model. The following device variables can be used:



<https://github.com/alanb128/opendatacam-ei>



balena dashboard | opendatacam ei

https://dashboard.balena-cloud.com/fleets/1951983

Getting Started Docs Forums Status Marc Pouss

Organizations

alanboris's Organization

Fleets

opendatacam-ei

Summary

Devices

Releases

Variables

Configuration

Settings

Provisioning Keys

Members

Teams

v15.0.5

Add new device

NANO

Select device type: Nvidia Jetson Nano 2GB Devkit SD

Select version: v2.98.12 (recommended)

Show outdated versions

Select edition: Development Recommended for first time users

Development images should be used when you are developing an application and want to use the fast local mode workflow. This variant should never be used in production.

Production

Production images are ready for production deployments, but don't offer easy access for local development.

Network Connection: WiFi + Ethernet

WiFi SSID:

WiFi Passphrase:

+ Advanced

Flash

Instructions

- 1 Use the form on the left to configure and download balenaOS for your new device.
- 2 Put the NVIDIA Jetson Nano 2GB Devkit in recovery mode
- 3 Unzip BalenaOS image and use Jetson Flash to provision the device.
- 4 After flashing of the Nano SD-CARD module is completed, please wait until the board is rebooted
- 5 Your device should appear in your fleet in the dashboard within a few minutes. Have fun!

Learn more

Get release

Views

IP address: 192.168.1.48 192.168.1.170

192.168.1.170

Help

For more details please refer to our [Getting Started Guide](#).

balena-os/jetson-flash: This tool allows users to flash BalenaOS on Jetson supported devices

https://github.com/balena-os/jetson-flash/

Search or jump to... Pull requests Issues Marketplace Explore

Code Issues 3 Pull requests 3 Actions Projects Wiki Security Insights

Code

master 28 branches 34 tags

Go to file Add file Code

acostach Merge pull request #88 from balena-os/nano_eMMC_update_32_7_2 ... b254209 11 days ago 140 commits

assets lib/resin-jetson-flash: Always boot from balenaOS supported mediums 4 months ago

bin jetson-tx2: Allow specifying odmdata value 8 months ago

lib lib: Update Jetson Nano SD-CARD to L4T 32.7.2 11 days ago

scripts flasher-unwrap: Increase progress events timeout 2 years ago

CHANGELOG.md v0.5.28 11 days ago

LICENSE Initial commit 5 years ago

README.md lib: Update Jetson Nano SD-CARD to L4T 32.7.2 11 days ago

package.json v0.5.28 11 days ago

About

This tool allows users to flash BalenaOS on Jetson supported devices

Readme Apache-2.0 license 32 stars 5 watching 9 forks

Releases 34 tags Create a new release

Packages No packages published Publish your first package

Contributors 7

jetson-flash

This tool allows users to flash BalenaOS on Jetson supported devices

This tool is separate into two parts:

- Extract BalenaOS image from a BalenaOS flasher image (this will be moved to etcher once the fatfs issues are fixed)
- Flash BalenaOS via USB on a Jetson board (this will be moved to etcher)

Balena devices support

- Jetson Nano eMMC - L4T 32.7.2
- Jetson Nano SD-CARD Devkit - L4T 32.7.2

Languages

JavaScript 100.0%

A screenshot of a macOS terminal window titled "-zsh". The window shows a file listing from the directory "jetson-flash" with the command "ls -la". The listing includes files like ".git", "CHANGELOG.md", "LICENSE", "README.md", "assets", "balena-cloud-jetson-nano-2GB-jetson-nano-2gb-devkit-2.98.12-v12.11.38.img", "bin", "lib", "package.json", and "scripts". Below the listing, the user runs the command "./bin/cmd.js -f balena-cloud-jetson-nano-2GB-jetson-nano-2gb-devkit-2.98.12-v12.11.38.img -m jetson-nano-2gb-devkit". At the bottom of the terminal, there is a progress bar indicating a download or upload operation: "226 kB↓" (blue), "449 kB↑" (red), and "28/9, 6:34 PM".

```
marcpous@Marc-macbook jetson-flash % ls -la
total 2670640
drwxr-xr-x 12 marcpous staff      384 28 set 00:33 .
drwxr-xr-x  7 marcpous staff      224 18 ago 15:31 ..
drwxr-xr-x 12 marcpous staff      384 21 jul 12:49 .git
-rw-r--r--  1 marcpous staff    3683 21 jul 12:49 CHANGELOG.md
-rw-r--r--  1 marcpous staff   11357 21 jul 12:49 LICENSE
-rw-r--r--  1 marcpous staff    3714 21 jul 12:49 README.md
drwxr-xr-x 12 marcpous staff      384 21 jul 12:49 assets
-rw-r--r--@  1 marcpous staff 1367343104 21 jul 09:07 balena-cloud-jetson-nano-2GB-jetson-nano-2gb-devkit-2.98.12-v12.11.38.img
drwxr-xr-x  3 marcpous staff      96 21 jul 12:49 bin
drwxr-xr-x  4 marcpous staff    128 21 jul 12:49 lib
-rw-r--r--  1 marcpous staff   1268 21 jul 12:49 package.json
drwxr-xr-x  3 marcpous staff     96 21 jul 12:49 scripts
marcpous@Marc-macbook jetson-flash %
marcpous@Marc-macbook jetson-flash %
marcpous@Marc-macbook jetson-flash %
marcpous@Marc-macbook jetson-flash % ./bin/cmd.js -f balena-cloud-jetson-nano-2GB-jetson-nano-2gb-devkit-2.98.12-v12.11.38.img -m jetson-nano-2gb-devkit
```

balena dashboard | opendatacam-ei +

https://dashboard.balena-cloud.com/fleets/1951983

90% ⚡ Getting Started Docs Forums Status Marc Pous 🌐

opendatacam-ei

Devices 2

Online Config Updating Offline Past prev Inactive

Releases 20

track latest

OpenDataCam Learn more

Create release

Summary

Devices Releases Variables Configuration Settings Provisioning Keys Members Teams

Add device Modify Search entries...

Name	Status	Device type	Last seen	Created on	UUID	OS version	OS line	OS variant	Supervisor version	IP address	⋮
alan-2gb	✓ Online	Nano Nvidia Jetson Nano 2GB Devkit SD	Online (for 2 hours)	Aug 17th 2022, 8:04 PM	61404a9	balenaOS 2.98.12	Production	12.11.38	192.168.1.170		⋮
Marc	✓ Online (VPN only)	Nano Nvidia Jetson Nano 2GB Devkit SD	Online (for a few seconds)	Jul 22nd 2022, 2:44 PM	7d309a9	balenaOS 2.98.12	Development	12.11.38	192.168.1.48	192.168.1.49	⋮

v15.0.6 <

Need help 🎁

balena dashboard | opendatacam

https://dashboard.balena-cloud.com/fleets/1951983

Getting Started Docs Forums Status Marc Pous

90%

Devices 2

Releases 20

track latest

OpenDataCam Learn more

Create release

Add device Modify Search entries...

Map Hybrid

v15.0.5

IMAGINE

EDGE IMPULSE

balena dashboard | Marc

https://dashboard.balena-cloud.com/devices/7d309a91696fd6ebbeaad905a7a324f6

90% ⚡ Getting Started Docs Forums Status Marc Pous 🌐

Marc

Status Online **UUID** 7d309a91696fd6ebbeaad905a7a324f6 **Type** Nvidia Jetson Nano 2GB Devkit SD

Online for 44 minutes **HOST OS VERSION** balenaOS 2.98.12 **SUPERVISOR VERSION** 12.11.38

CURRENT RELEASE 4c3418b **TARGET RELEASE** 4c3418b

LOCAL IP ADDRESS 192.168.1.48 192.168.1.49 **PUBLIC IP ADDRESS** 79.153.124.126 **MAC ADDRESS** 48:B0:2D:2E:5C:A3
00:E0:4C:49:D5:12

TAGS (0) No tags configured yet

NOTES
<http://www.insecam.org/en/bytag/City/> http://85.158.74.11/mjpg/video.mjpg http://75.162.195.79:80/jpg/image.jpg?1660759872
<http://84.118.176.178:81/mjpg/video.mjpg> http://153.232.32.204:8080/?action=stream http://46.165.133.142/image

Services

Service	Status	Release
audio	Running	4c3418b
ei	Running	4c3418b
mongo	Running	4c3418b
opendatacam	Running	4c3418b

CPU ~42% **Temperature** ~45C **Memory** 1000 MB/1.9 GB **Storage** (dev/mmcblk0p1) 7.7 GB/28.0 GB

Logs

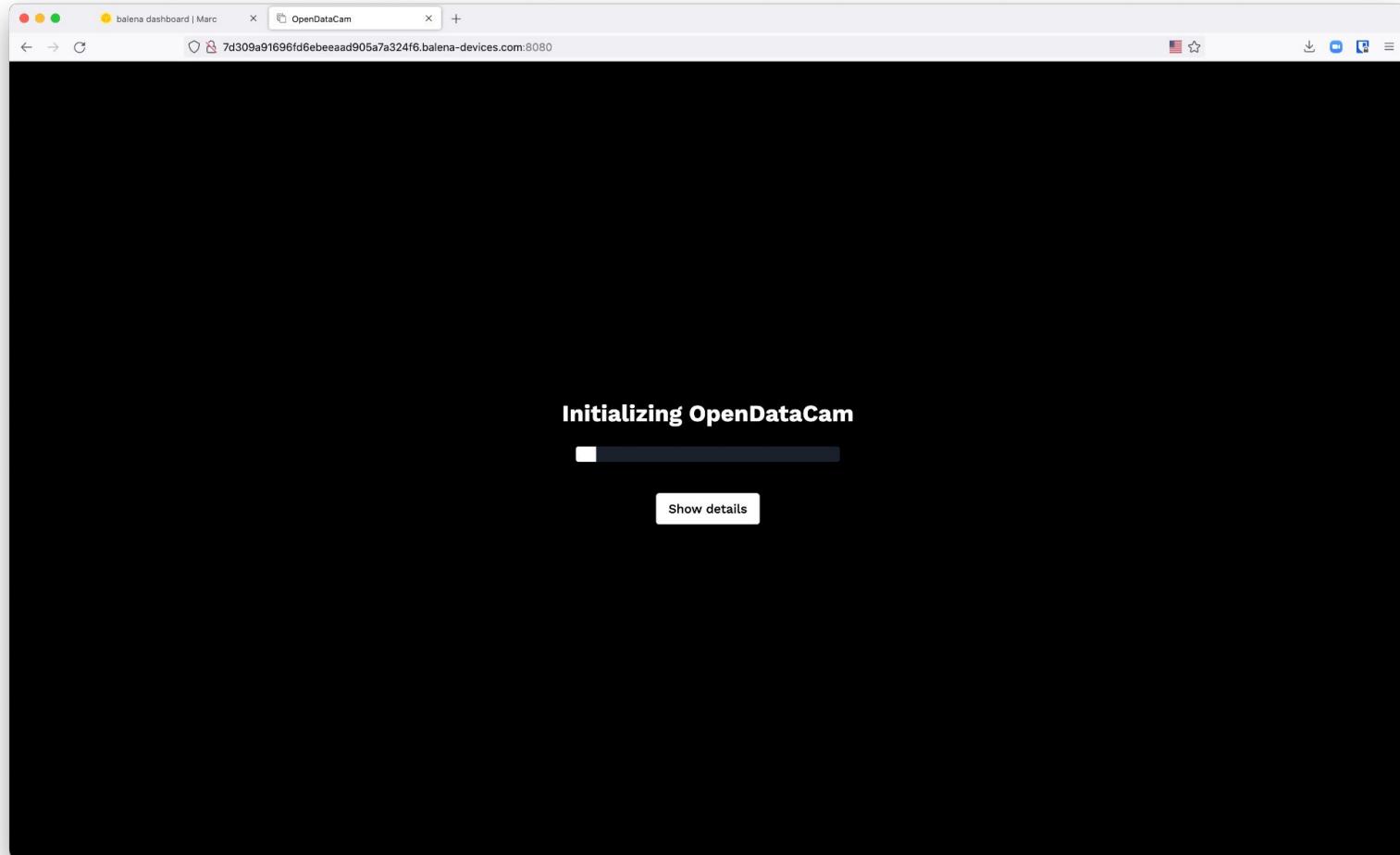
Search entries... Add filter Views UTC Timestamps

```
gunment: "device_id="0" name="platform-70030000.hda" card_name="alsa_card.platform-70030000.hda" failIfFalse tsched=yes fixed_latency_range=no ignore_dbe=no deferred_volume=yes use_ucm=yes avoid_resampling=yes card_properties="module-udev-detect.discovered=1"**: initialization failed.  
27.09.22 22:03:56 (+0000) ei Edge Impulse Linux runner v1.3.9  
27.09.22 22:03:57 (+0000) ei [RUN] Downloading model...  
27.09.22 22:03:59 (+0000) ei [RUN] Downloading model OK  
27.09.22 22:03:59 (+0000) ei [RUN] Stored model in /usr/src/app/modelfile.eim  
27.09.22 22:07:24 (+0000) mongo 2022-09-27T22:07:24.479+0000 I NETWORK [conn2] end connection 172.18.0.5:46156 (2 connections now open)  
27.09.22 22:07:24 (+0000) mongo 2022-09-27T22:07:24.482+0000 I NETWORK [conn3] end connection 172.18.0.5:46156 (1 connection now open)
```

Terminal

Select a target Start terminal session

v15.0.5 Need help





balena dashboard | Marc traffic-22V - Dashboard - Edge

https://studio.edgeimpulse.com/studio/139317/

Edge Impulse Imagine! Join us for the latest innovations in edge machine learning for the real world, Sept 28-30. [Learn more.](#)

Alan / traffic-22V

Project info Keys Export

Alan / traffic-22V

This is your Edge Impulse project. From here you acquire new training data, design impulses and train models.

+ New tag

About this project

Add README

Creating your first impulse (100% complete)

Acquire data
Every Machine Learning project starts with data. You can capture data from a development board or your phone, or import data you already collected.
[LET'S COLLECT SOME DATA](#)

Design an Impulse
Teach the model to interpret previously unseen data, based on historical data. Use this to categorize new data, or to find anomalies in sensor readings.
[GETTING STARTED: CONTINUOUS MOTION RECOGNITION](#)
[GETTING STARTED: RESPONDING TO YOUR VOICE](#)
[GETTING STARTED: ADDING SIGHT TO YOUR SENSORS](#)

Sharing

Your project is private.

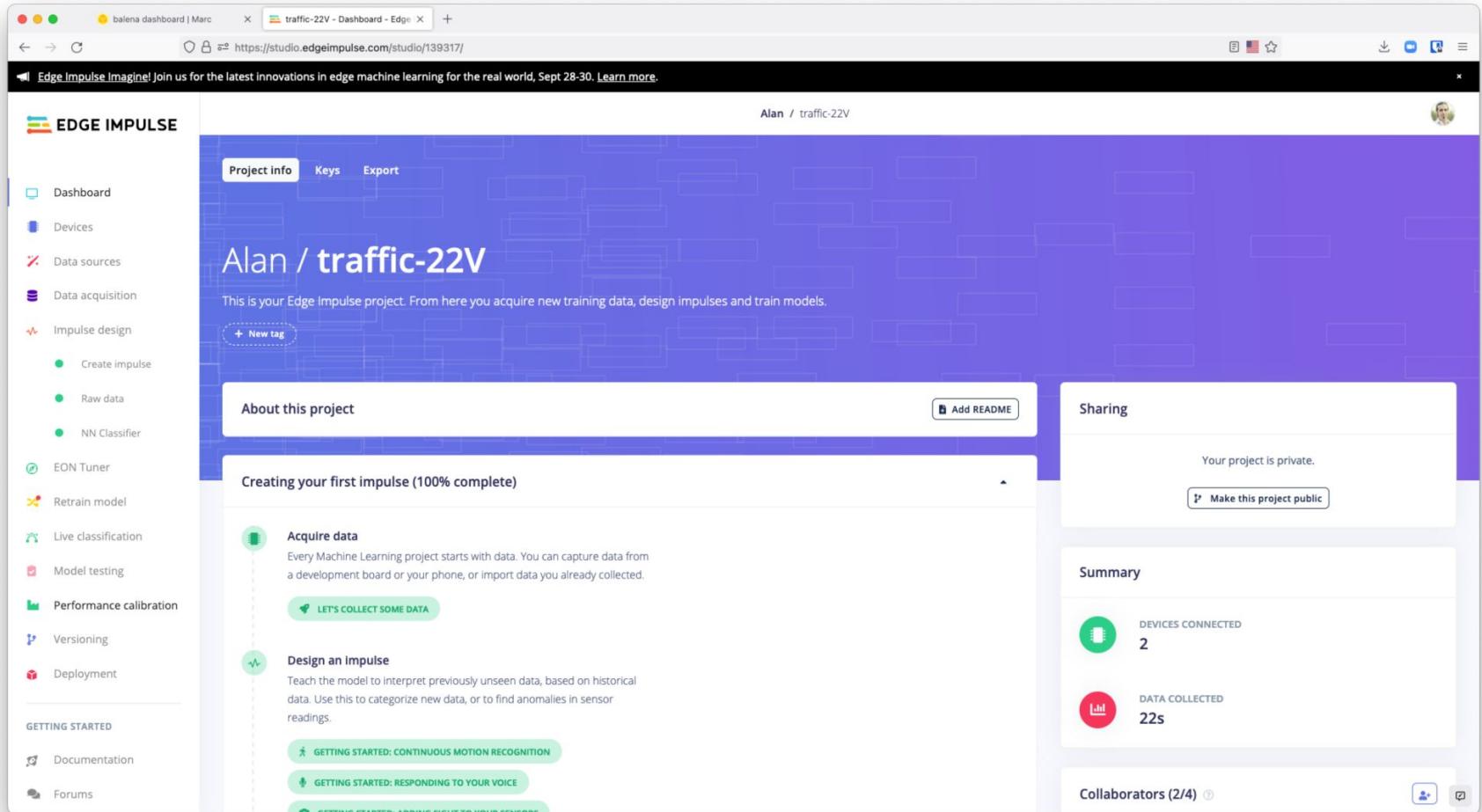
[Make this project public](#)

Summary

DEVICES CONNECTED
2

DATA COLLECTED
22s

Collaborators (2/4) [Add collaborator](#)



balena dashboard | Marc traffic-22V - Keys - Edge Impulse +

https://studio.edgeimpulse.com/studio/139317/keys

Edge Impulse Imagine! Join us for the latest innovations in edge machine learning for the real world, Sept 28-30. [Learn more.](#)

Alan / traffic-22V

EDGE IMPULSE

Project info Keys Export

API Keys

+ Add new API key

API keys are used to connect to the Edge Impulse API, and to connect devices to the Edge Impulse remote management API. All API keys that are not marked as development keys are truncated.

NAME	API KEY	ROLE	CREATED	DEVELOPMENT KEY	⋮
nvidia-jetson-22V-key	ei_a9eff7f54d2e882e8c8580f5be5f4095ef3453e7eec0d7489aa700ea15...	Admin	Sep 20 2022, 20:39:52	ⓘ	⋮

HMAC Keys

+ Add new HMAC key

HMAC keys are used to sign device data that gets sent through the [ingestion API](#). The data acquisition format documentation has examples on how to sign data in a variety of languages.

NAME	HMAC KEY	CREATED	DEVELOPMENT KEY	⋮
nvidia-jetson-22V-key	865ef88f6c5eb99e1503e6f70b21d937	Sep 20 2022, 20:39:52	ⓘ	⋮

© 2022 [Edgimpulse Inc.](#) All rights reserved

Dashboard Devices Data sources Data acquisition Impulse design Create impulse Raw data NN Classifier EON Tuner Retrain model Live classification Model testing Performance calibration Versioning Deployment

GETTING STARTED Documentation Forums

balena dashboard | alian-2gb

https://dashboard.balena-cloud.com/devices/61404a965d28cd804604376aee82d054/envvars

90% Getting Started Docs Forums Status Marc Pous 

Add variable Search entries...

Name	Fleet value	Device value	Service name	Actions
API_KEY	not defined	ei_69aa3d34a39f92c6e30a5081e5b9a4b48e4cef1...	ei	 
EI_API_KEY	not defined	ei_69aa3d34a39f92c6e30a5081e5b9a4b48e4cef1...	ei	 
HMAC_KEY	not defined	6f83e477e734cf1d58fcba6817e8ac99	ei	 
INPUT_REMOTE_CAM	YOUR IP CAM STREAM	rtsp://192.168.1.70/0	All services	 
INPUT_USBCAM	!v4l2src device=/dev/video0 ! video/x-raw,...		All services	
LINE_LEFT_NAME	not defined	line_south	ei	 
LINE_RIGHT_NAME	not defined	line_north	ei	 
REMOTE_CAM	not defined	usbcam	All services	 
UPLOAD_DATA_FILES	not defined	1	ei	 
VIDEO_INPUT	usbcam	remote_cam	All services	 

1 - 10 of 10 < >

v15.0.5

Need help 

balena dashboard | Marc traffic-22V - Data acquisition

https://studio.edgeimpulse.com/studio/139317/acquisition/training?page=1

Edge Impulse Imagine! Join us for the latest innovations in edge machine learning for the real world, Sept 28-30. [Learn more](#).

Alan / traffic-22V

EDGE IMPULSE

Training data Test data | Data explorer | Upload data Export data

DATA COLLECTED 18s

TRAIN / TEST SPLIT 82% / 18% ▲

Record new data

No devices connected to the remote management API.

RAW DATA standstill.sample0919200357

SAMPLE NAME	LABEL	ADDED	LENGTH
light.3dt3dg0f	light	Yesterday, 23:48:45	1s
light.3dt35us7	light	Yesterday, 23:44:38	1s
light.3dt2kqp3	light	Yesterday, 23:35:17	1s
light.3dt05707	light	Yesterday, 22:51:48	1s
standstill.sample0919200357	standstill	Yesterday, 04:34:19	1s
standstill.sample0919200257	standstill	Yesterday, 04:34:19	1s
standstill.sample0919195055	standstill	Yesterday, 04:34:18	1s
standstill.sample0919191951	standstill	Yesterday, 04:34:18	1s
standstill.sample0919191851	standstill	Yesterday, 04:34:18	1s
moderate.sample0919193253	moderate	Yesterday, 04:34:18	1s
moderate.sample0919193053	moderate	Yesterday, 04:34:18	1s
moderate.sample0919192552	moderate	Yesterday, 04:34:18	1s

car_count avg_speed

balena dashboard | Marc traffic-22V - Create impulse

https://studio.edgeimpulse.com/studio/139317/create-impulse

Edge Impulse Imagine! Join us for the latest innovations in edge machine learning for the real world, Sept 28-30. [Learn more.](#)

Alan / traffic-22V

EDGE IMPULSE

An impulse takes raw data, uses signal processing to extract features, and then uses a learning block to classify new data.

Dashboard Devices Data sources Data acquisition Impulse design Create impulse Raw data NN Classifier EON Tuner Retrain model Live classification Model testing Performance calibration Versioning Deployment

Time series data

Input axes (2)
car_count, avg_speed

Window size
1000 ms.

Window increase
1000 ms.

Frequency (Hz)
1

Zero-pad data

Raw Data

Name Raw data

Input axes (2)
 car_count
 avg_speed

Classification (Keras)

Name NN Classifier

Input features
 Raw data

Output features
3 (light, moderate, standstill)

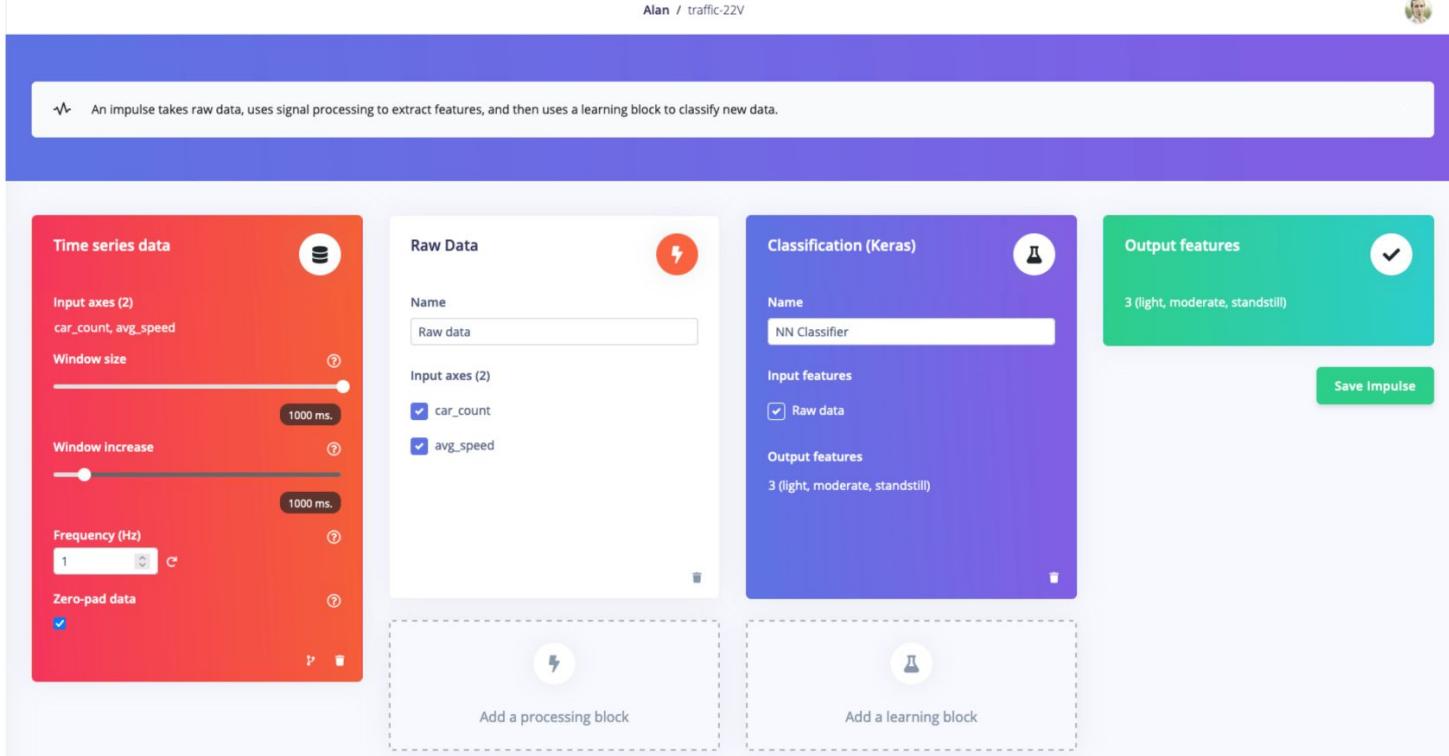
Output features

3 (light, moderate, standstill)

Save Impulse

Add a processing block

Add a learning block



balena dashboard | Marc traffic-22V - NN Classifier - Edi

https://studio.edgeimpulse.com/studio/139317/learning/keras/29

Alan / traffic-22V 

EDGE IMPULSE

#1 Click to set a description for this version

Neural Network settings

Training settings

Number of training cycles Learning rate Validation set size % Auto-balance dataset

Neural network architecture

Input layer (2 features)

Dense layer (20 neurons)

Dense layer (10 neurons)

Add an extra layer

Output layer (3 classes)

Start training

Training output

Model

Model version: Quantized (int8)

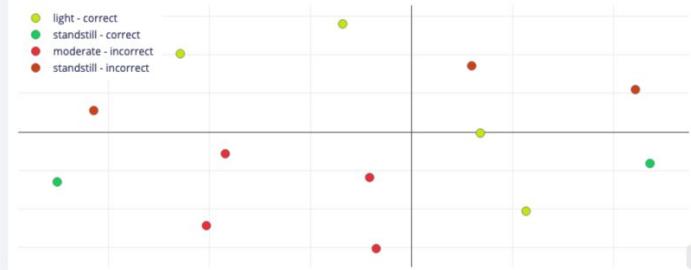
Last training performance (validation set)

ACCURACY 33.3% LOSS 11.76

Confusion matrix (validation set)

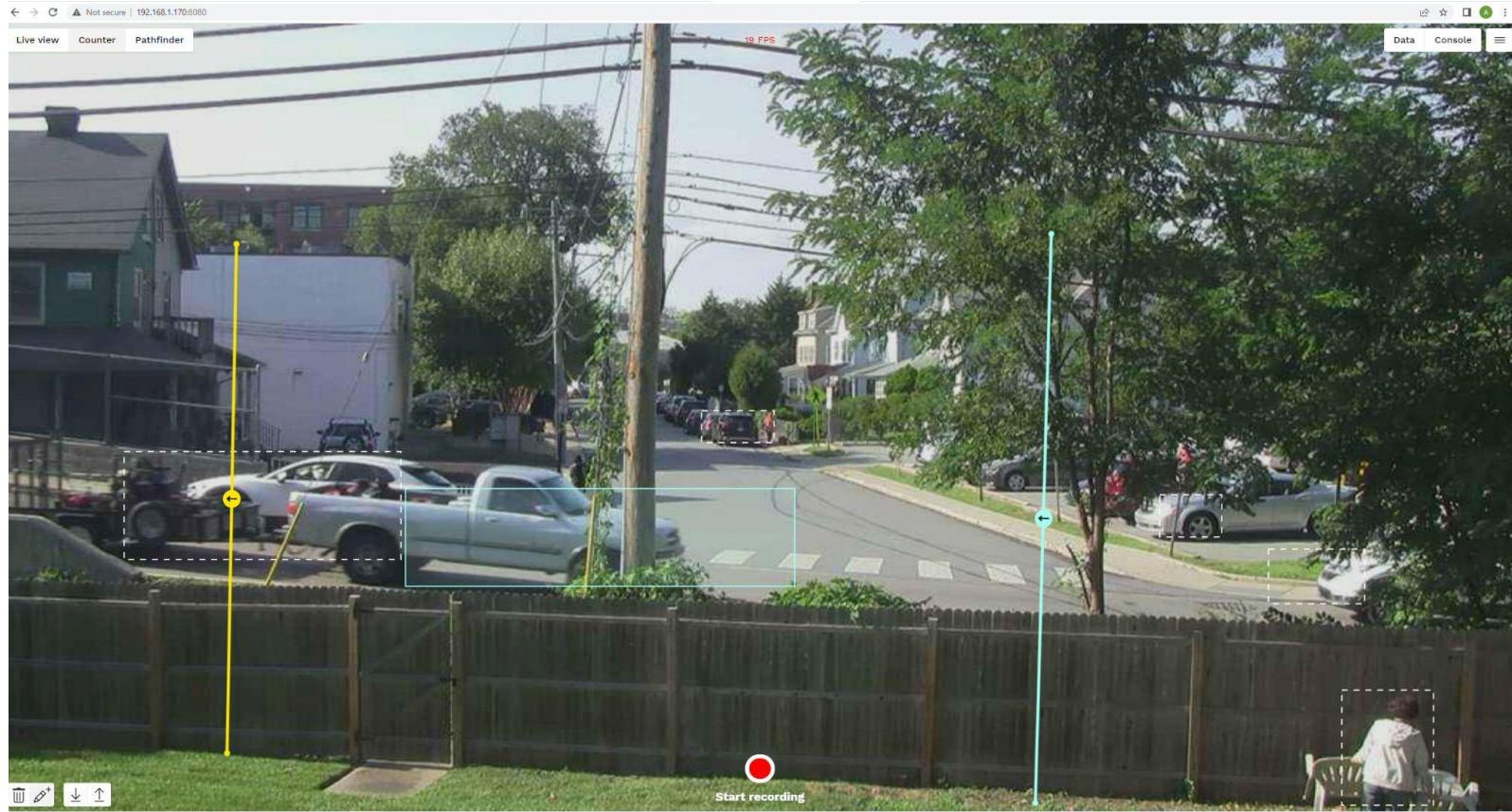
	LIGHT	Moderate	STANDSTILL
LIGHT	~	~	~
Moderate	0%	0%	100%
STANDSTILL	50%	0%	50%
F1 SCORE	0.00	0.00	0.50

Data explorer (full training set)



Legend:

- light - correct
- standstill - correct
- moderate - incorrect
- standstill - incorrect



```
#####
Started recording #1 for next 60 second(s)...
Response from odc: OK
45.0 s. remaining...
30.0 s. remaining...
15.0 s. remaining...
Stopping recording...
Found 2 car journey(s).
Final sample data: [2, 10138.0]
SAVE_DATA_FILES=True; Saved to file: light.sample0928194333.csv

Starting local EI inference on model...
Loaded runner for "Alan / traffic-22V"
^^^^ Local classification results: ^^^^
moderate: 0%
standstill: 1%
light: 98%
Timing information: {'anomaly': 0, 'classification': 0, 'dsp': 0, 'json': 0, 'stdin': 9}

Preparing to upload data file to Edge Impulse
Type a label or enter for [light]
light
Uploaded file to Edge Impulse 200 b'light.3dveni5p.json'

#####
Started recording #2 for next 60 second(s)...
```

Discuss why the model only has 33%
accuracy but this is on purpose!

This is an example to push developers to
apply innovation using these type of
systems.

**Let's update
the fleet.**

**Here, live,
remotely!**



balena dashboard | alan-2gb opendatacam-ei/ei_run.sh at master · alansb128/opendatacam-ei · GitHub

https://github.com/alansb128/opendatacam-ei/blob/main/ei/ei_run.sh

Search or jump to... Pull requests Issues Marketplace Explore

alanb128 / opendatacam-ei Public

Code Issues Pull requests Actions Projects Wiki Security Insights

main · opendatacam-ei / ei / ei_run.sh Go to file ...

alanb128 Update ei_run.sh Latest commit fa1434f yesterday History

1 contributor

16 lines (11 sloc) | 497 Bytes Raw Blame

```
1 #!/bin/sh
2
3 # Here we download the latest model each time the container starts
4
5 # EI_API_KEY defined as environment variables in BalenaCloud
6 edge-impulse-linux-runner --api-key $EI_API_KEY --download modelfile.eim
7
8 # Give the downloader some breathing time
9 sleep 10
10
11 # To start our Python script that pulls data from Opendatacam,
12 # runs inferences, and optionally sends training files back to EI cloud,
13 # uncomment the line below and comment the sleep line.
14 # python3 runner.py
15
16 sleep infinity
```

© 2022 GitHub, Inc. Terms Privacy Security Status Docs Contact GitHub Pricing API Training Blog About

Summary

Takeaways

- Fleet owners and ML owners need balena.io and Edge Impulse.
- Separate edge inferencing stack on CPU and GPU using the full potential of devices.
- Start using 21st century development tools (containers, CI/CD, open-source and more).



Alan Boris

Hardware Hacker at balena.io

alan@balena.io

<https://github.com/alanb128/>

@balena_io

Contribute, share lessons
learned and ask questions

Running AI models on the NVIDIA Jetsons CPU and GPU with balena



Alan Boris and Marc Pous

balena.io

Solutions in this workshop for

- Deploying an AI model to a balena fleet of Jetson devices.
- Remotely changing device variables.
- Separate edge inferencing stack on CPU and GPU.
- Sending sample data back to EI for re-training.
- Pushing an updated model back out to a fleet.

Demo:

- Run separate, local AI inference on GPU and CPU
- Feed
- Set device variable to collect data, label and send to EI
- Re-train on EI, then start inferencing with new model
- Demonstrate how we could use a completely different model by setting a device variable