

ET02 DATE24 Tutorial

On-Device Continual Learning

Meets Ultra-Low Power Processing



Manuele Rusci, KU Leuven
manuele.rusci@kuleuven.be



Davide Nadalini, Università di Bologna
d.nadalini@unibo.it



Cristian Cioflan, ETH Zürich
cioflanc@iis.ee.ethz.ch

Content:

[Part I] Introduction to **On-Device Continual Learning** (M. Rusci) [15']

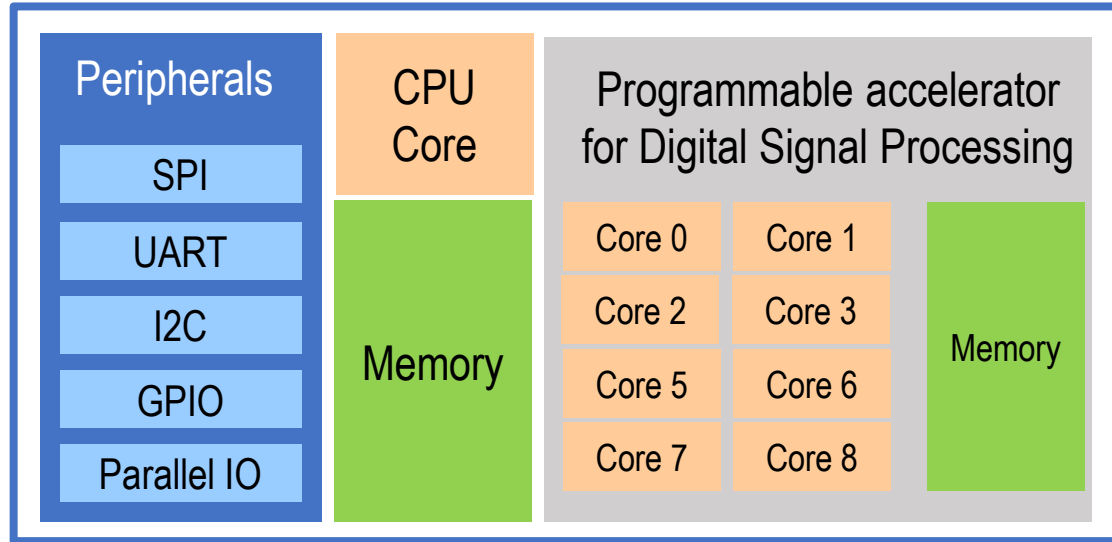
[Part II] DNN Training Primitives and Mapping to Microcontrollers (MCUs) (C. Cioflan) [30']

[Part III] PulpTrain-Lib Hands-on (D. Nadalini) [30']

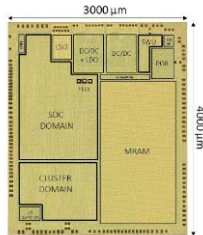
[Part IV] Trends and Tricks for Efficient On-Device Continual Learning (M. Rusci) [10']



Microcontrollers (MCUs)



- Ultra-low power near-sensor processing:
 - Less than tens of mW
- Heterogeneous platforms for low-latency processing but still flexible!!
 - Software-programmable
- PULP platform architectural template:
 - Open-source RISC-V based MCU
 - Multi-core cluster with tightly-coupled memory
- Hands-on using the software simulator GVOSC

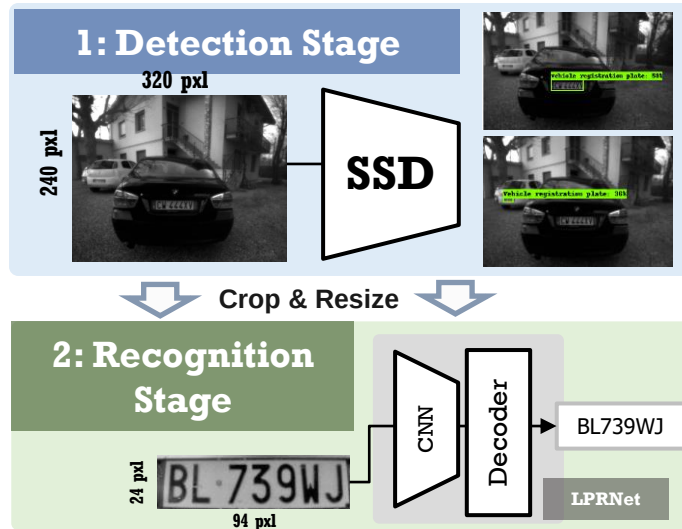


VEGA
[Rossi]



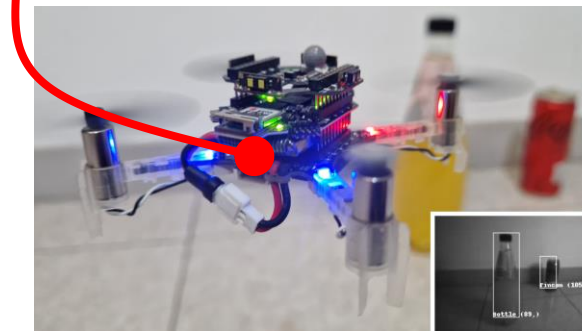
GAP9
[Greenwaves
Technologies]

Deep Learning at the Extreme Edge

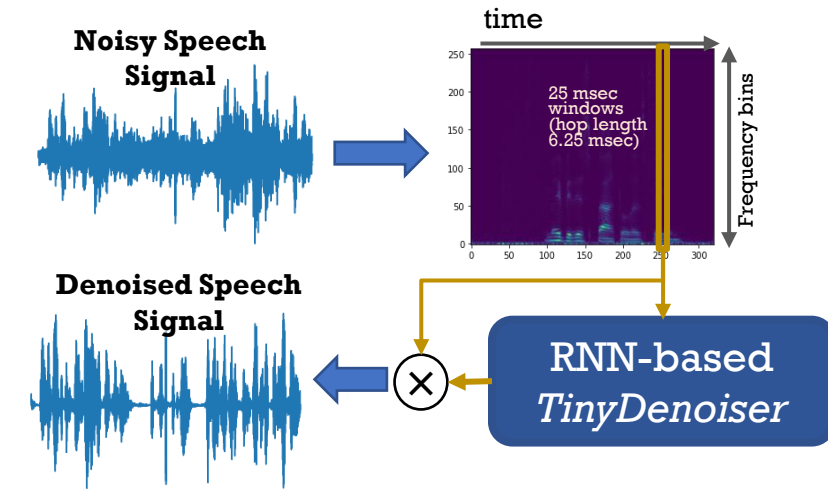


[Lamberti2021]

On-board processing
on a nano-drone

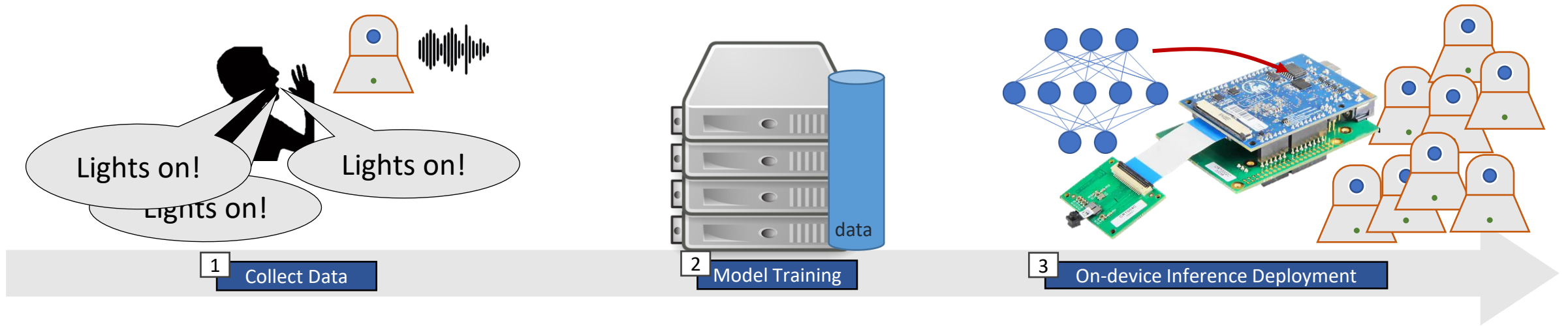


[Lamberti2023]

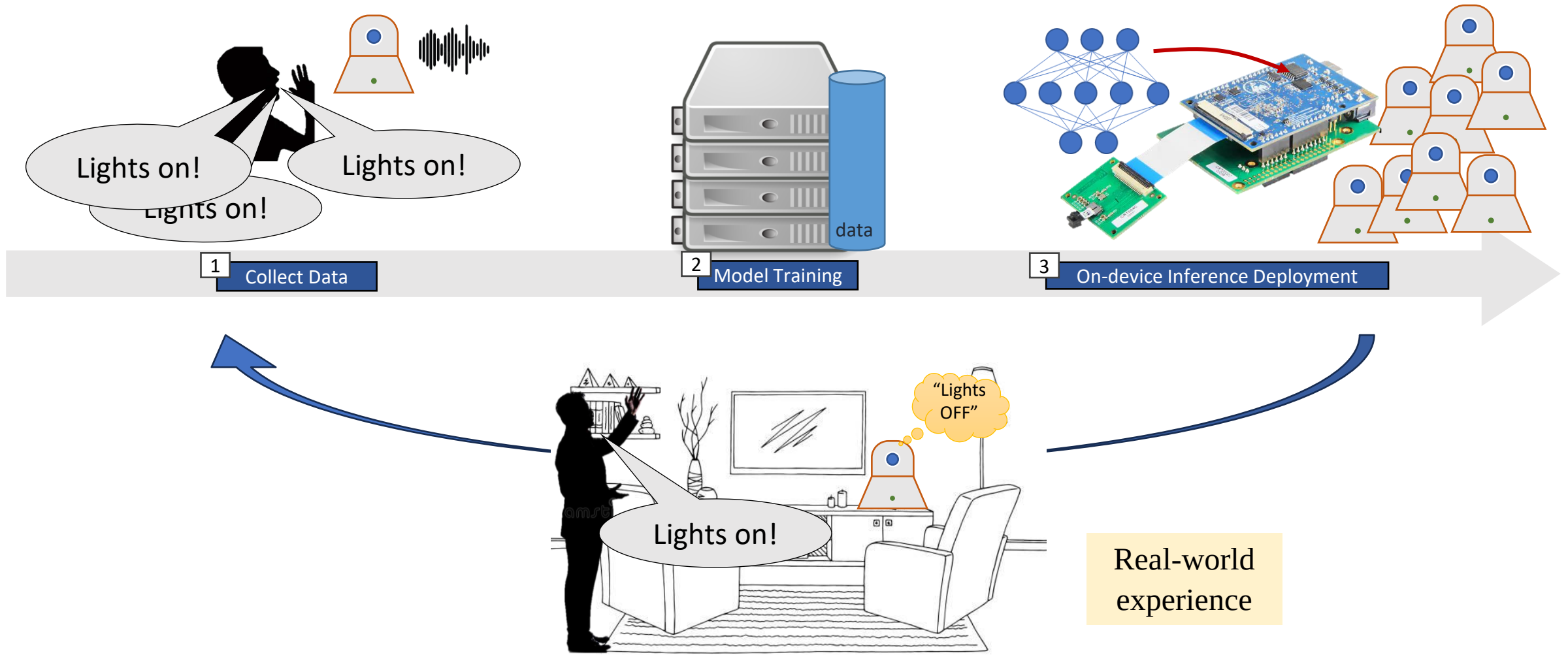


[Rusci2022]

Train-Once-Deploy-Everywhere



Train-Once-Deploy-Everywhere



Real-world *Failures*

- Generalization to unknown environments
 - original data $\neq$ target data
 - Model capacity and shortcuts
- Domain shifts over time
- Sensor-specific data
- **Personalization**
 - E.g. new objects, new keywords (new classes)



Training Data
(Google Open Images)



Real Data
(Low-Power Camera)

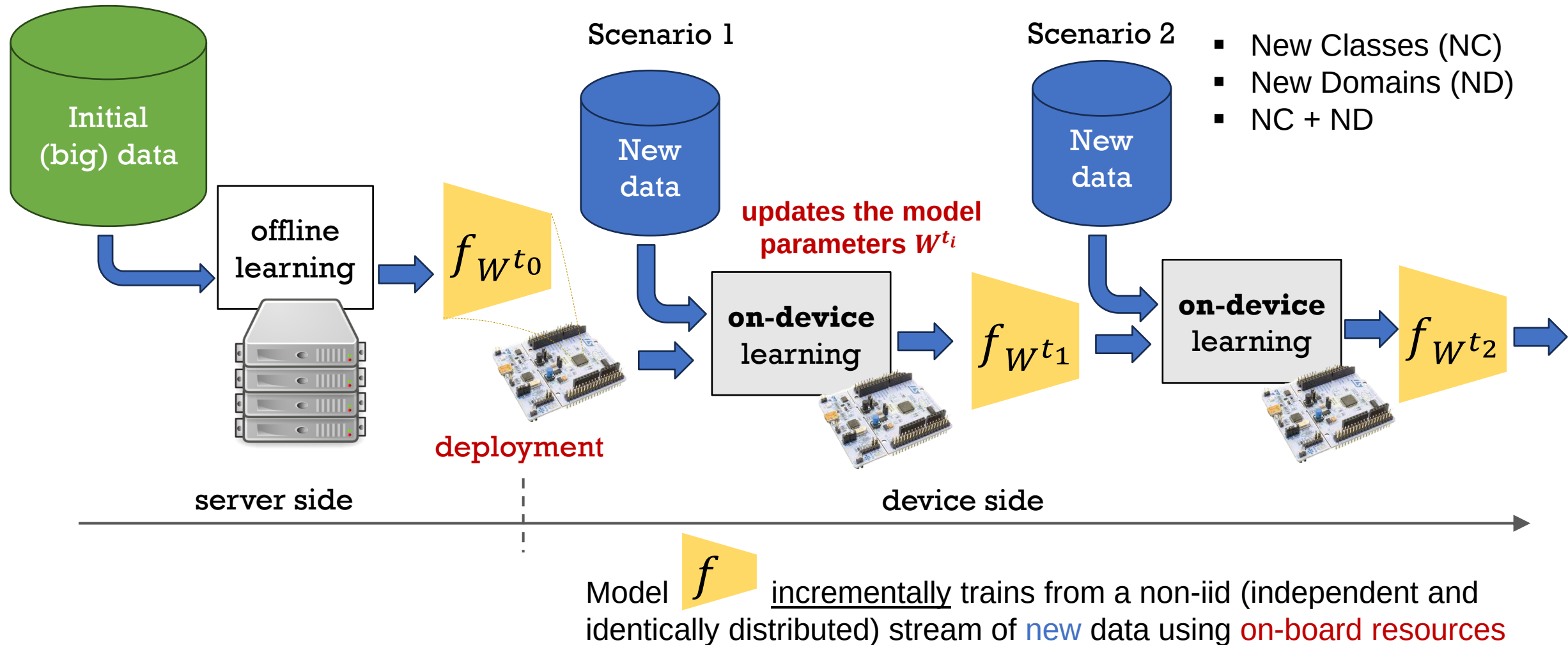


Obj Detector trained on
Google Open Images

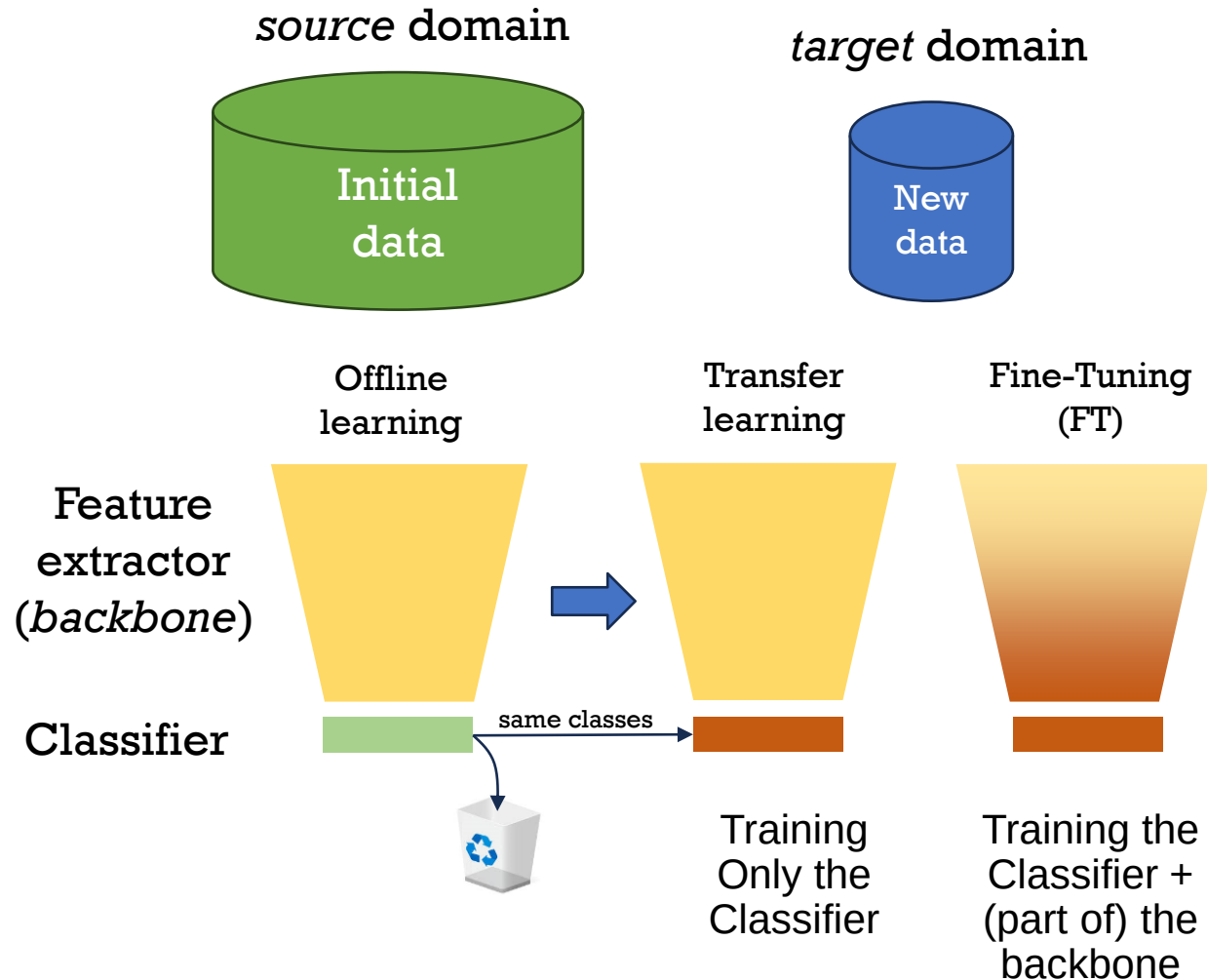


After finetuning on
Low-Power Camera data

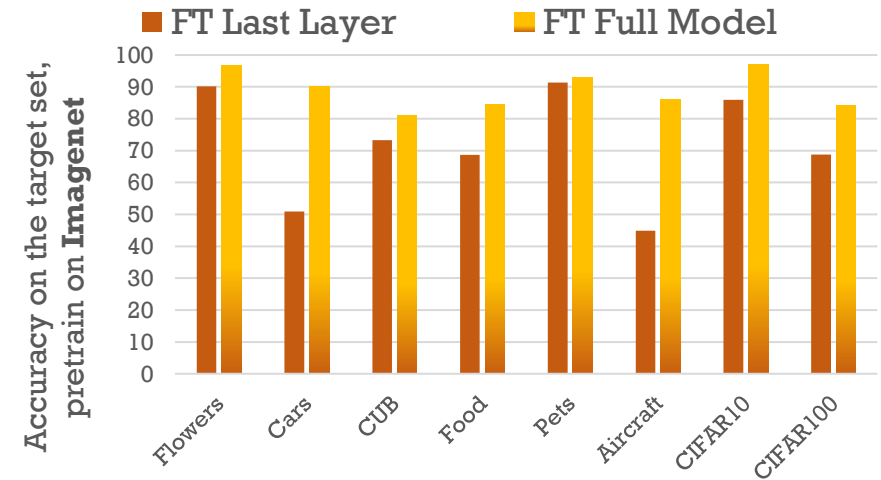
Continual **On-Device** Learning (ODL)



Transfer-Learning and Fine-Tuning



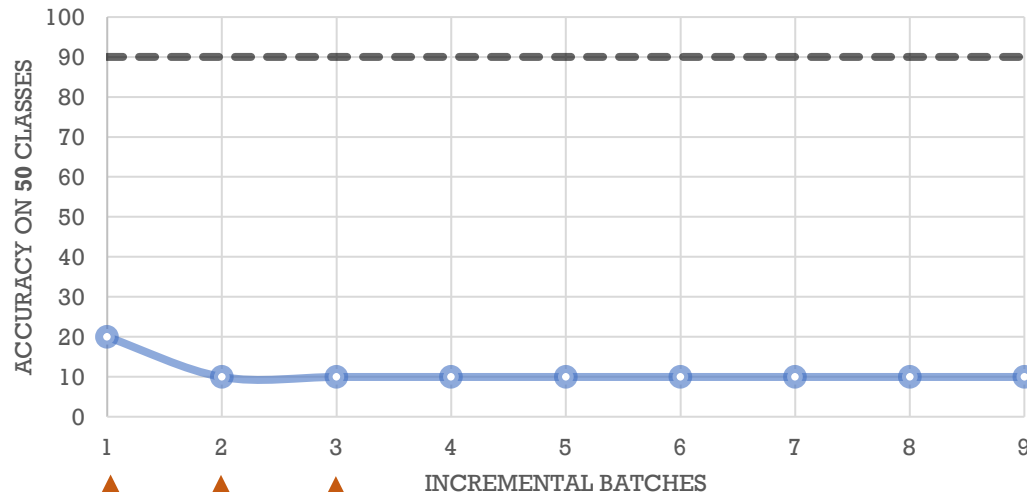
Adaptation to the target domain using FT [Cai2020]



Catastrophic Forgetting

Class-incremental Learning scenario

[Maltoni2019]



Joint Training: model trained not continually on the full dataset (upper bound)

Fine-tuning: training only on samples of current tasks

Catastrophic Forgetting!!

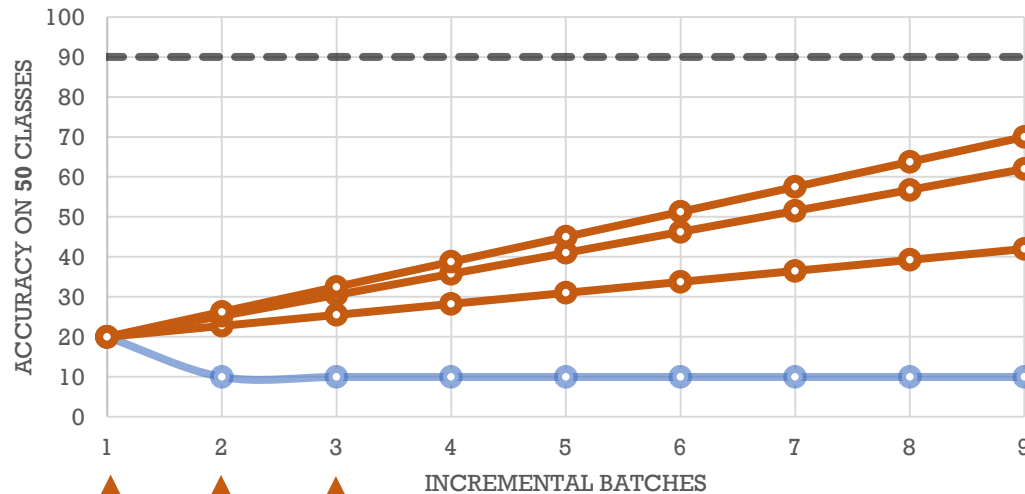
Core50 [Lomonaco2017]

- First task of 10 classes
- 8 incremental sessions of 5 tasks

Catastrophic Forgetting

Class-incremental Learning scenario

[Maltoni2019]



Core50 [Lomonaco2017]

- First task of 10 classes
- 8 incremental sessions of 5 tasks

Joint Training: model trained not continually on the full dataset (upper bound)

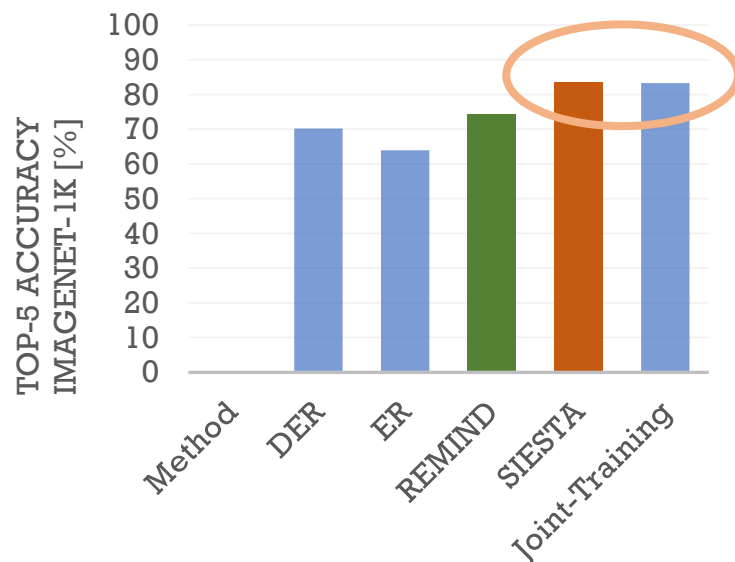
Fine-tuning: training only on samples of current tasks

Catastrophic Forgetting!!

Continual Learning (CL) strategies to deal with catastrophic forgetting [vandeVen2022, DeLange2021]

CL reached offline learning accuracy

SIESTA [Harun2023] showed same Top5 Imagenet-1k accuracy of offline learner but class-incrementally trained (data streams of 100 classes)

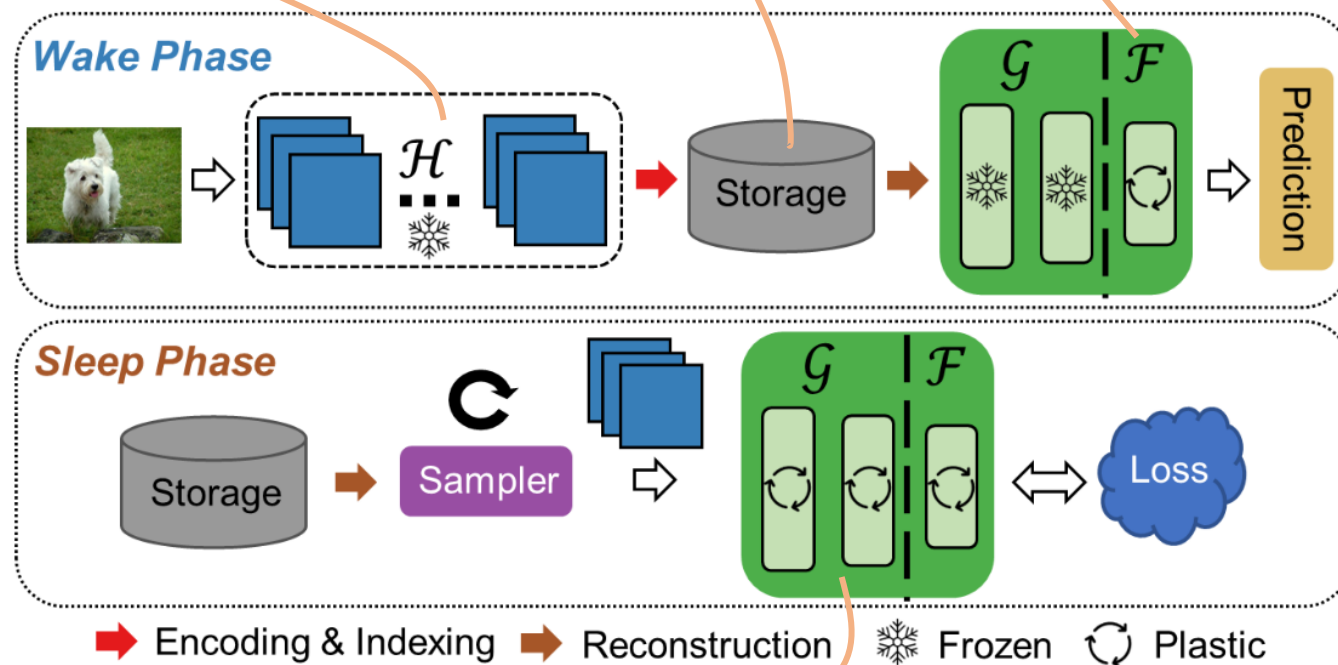


trained on first 100 classes and then frozen (self-supervised init)

Storing (part of) of *old data* in a compressed form

Online updates on the last layer

$$\mathbf{f}_k \leftarrow \frac{c_k \mathbf{f}_k + \mathbf{z}_t}{c_k + 1}$$



Batch **learning** of $\mathcal{G}$ and $\mathcal{F}$ using rehearsal

Can we bring Continual Learning on Ultra Low Power MCUs?

- Deep Neural Network Training explained (C. Cioflan)
 - Basic Operations of Backpropagation-based training
 - Mapping to MCUs and memory compute and costs
 - This tutorial: supervised learning, no spiking models, no continual learning algorithms
- Hands-on practice: training on a multi-core MCUs (D.Nadalini)
- Recent Trends on On-Device Learning (M. Rusci)

Reference

- [Lamberti2021] Lamberti, Lorenzo, Manuele Rusci, Marco Fariselli, Francesco Paci, and Luca Benini. "Low-power license plate detection and recognition on a risc-v multi-core mcu-based vision system." ISCAS, 2021.
- [Lamberti2023] Lamberti, Lorenzo, Luca Bompani, Victor Javier Kartsch, Manuele Rusci, Daniele Palossi, and Luca Benini. "Bio-inspired autonomous exploration policies with cnn-based object detection on nano-drones." DATE, 2023.
- [Rusci2022] Rusci, Manuele, Marco Fariselli, Martin Croome, Francesco Paci, and Eric Flamand. "Accelerating RNN-Based Speech Enhancement on a Multi-core MCU with Mixed FP16-INT8 Post-training Quantization." In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 2022.
- [Cai2020] Cai, Han, Chuang Gan, Ligeng Zhu, and Song Han. "Tinytl: Reduce activations, not trainable parameters for efficient on-device learning." *Neurips* 2020.
- [DeLange2021] De Lange, Matthias, Rahaf Aljundi, Marc Masana, Sarah Parisot, Xu Jia, Aleš Leonardis, Gregory Slabaugh, and Tinne Tuytelaars. "A continual learning survey: Defying forgetting in classification tasks." *IEEE TPAMI* 44, 2021.
- [vandeVen2022] van de Ven, Gido M., Tinne Tuytelaars, and Andreas S. Tolias. "Three types of incremental learning." *Nature Machine Intelligence* 4, 2022.
- [Lomonaco2017] V. Lomonaco "COrE50: a new Dataset and Benchmark for Continuous Object Recognition". CoRL 2017.
- [Maltoni2019] Maltoni, Davide, and Vincenzo Lomonaco. "Continuous learning in single-incremental-task scenarios." *Neural Networks* 116 (2019): 56-73.
- [Harun2023] Harun, et al. "SIESTA: Efficient Online Continual Learning with Sleep." *TLMR* 2023.



Questions

manuele.rusci@esat.kuleuven.be