On-Device Federated Continual Learning on RISC-V-based Ultra-Low-Power SoC for Intelligent Nano-Drone Swarms

Lars Kröger†, **Cristian Cioflan‡**, Victor Kartsch‡, Luca Benini‡§

†D-ITET, ETH Zürich, Switzerland, ‡IIS, ETH Zürich, Switzerland, §DEI, University of Bologna, Italy

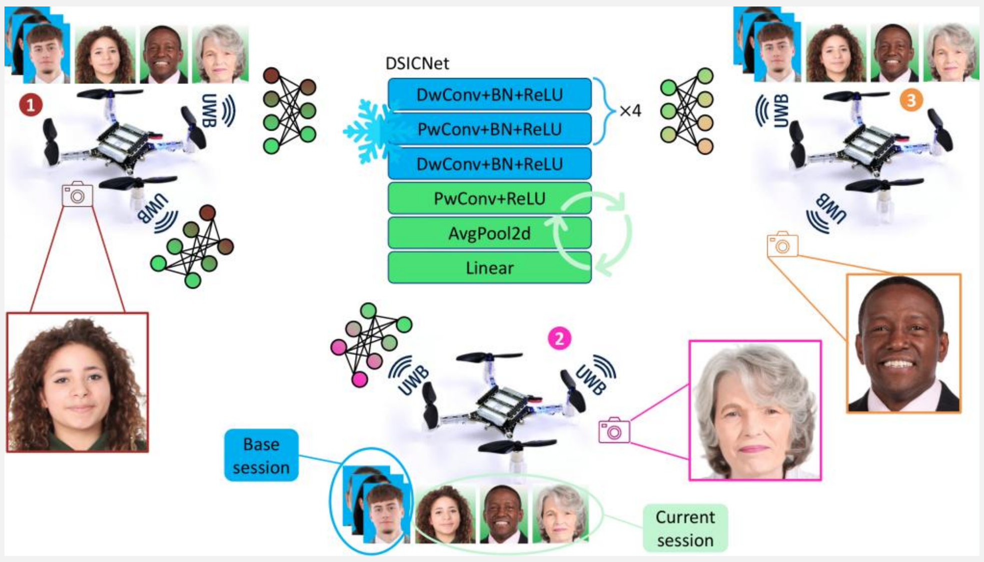
1 Introduction and Motivation

Continual Learning: on-site samples and offline data belong to different distributions. Intelligent models must **expand** their **knowledge domain** and **remember past experiences**.

Federated Learning: intelligent agents must **share and build knowledge** without exchanging sensitive information.

On-Device Learning: Ultra-low-power SoCs have reduced **on-chip memory, computational resources,** and **device lifetime**. The extreme edge hardware-associated costs must be **addressed without sacrificing recognition accuracy**.

2 ODFCL System



Mean Output Loss: regularize past vs. novel classes

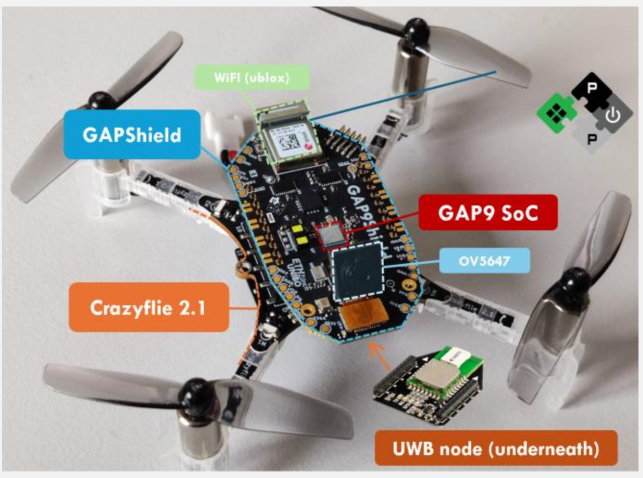
$$L_{MOL} \triangleq \left|\left|\frac{1}{N_{base}}\sum_{i \in N_{base}} y_i - \frac{1}{N_{current}-1}\sum_{i \in \frac{N_{current}}{\{gt\}}} y_i\right|\right|^2$$

FedProx [1]: regularize local vs. global model

$$L_{FedProx} \triangleq \left|\left|w_{local} - \frac{1}{nodes}\sum_{n \in nodes} w_{n_{local}}\right|\right|^2$$

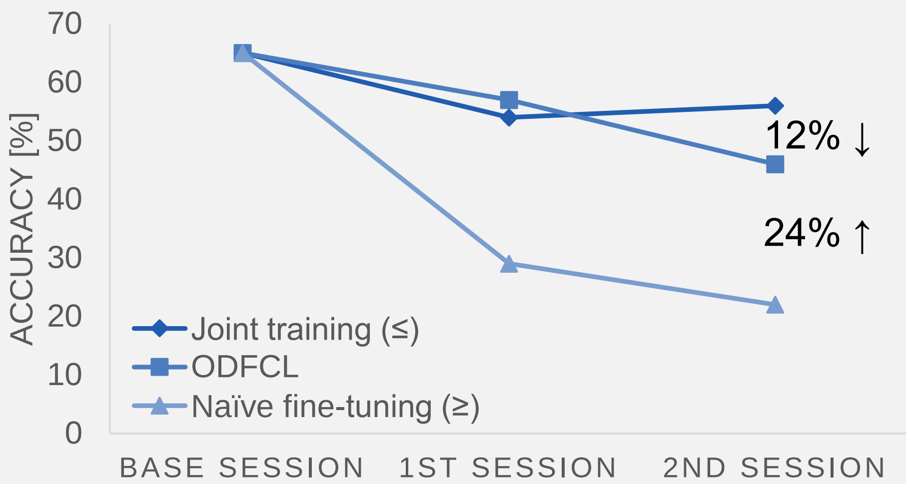
DSICNet [2] with 30 kparameters:

- **int8-quantized** backbone;
- **fp32** layers trainable via backpropagation.



GAP9Shield[3] integrating the GAP9 SoC with 10 RISC-V cores. The shield and the UWB Loco Positioning Deck [4] are mounted on the Craziflie 2.1 nano-drone [5].

3 Results



DSICNet is pretrained on CIFAR100. The **base session: learning** four faces from LFW [6]. **1st** and **2nd sessions: continually learning** three new faces per session, each nano-drone node exposed to one face per session.

We need only **28 samples/class** to recognize 10 faces with 46% accuracy, 2x better than naïve fine-tuning, only 12% below an ideal offline scenario.

The **local training latency** is only **178 ms**, yet it takes the nodes **1.7 s to exchange the weights**, for a total of **10.5 s per epoch**.

Operating conditions	**f = 240 MHz V = 650 mV**	**f = 370 MHz V = 800 mV**
Latency [ms]	178.4	117.6
Power [mW]	24.4	53.1
Energy [mJ]	4.3	6.2
ODL peak memory [kB]	29	
FL memory [kB]	24	

4 Conclusion

ODFCL paves the way towards multi-agent lifelong learning at the extreme edge. Ongoing work includes:

- deploying larger networks (exhausting the remaining 400 kB of RAM) to improve the baseline accuracy;
- improving the continual learning strategy to reduce forgetting;
- optimizing the communication protocol to improve latency.

References

[1] Tian Li et al. “Federated Optimization in Heterogeneous Networks”. In: Proc. of Machine Learning and Systems. 2020, pp. 429–450.

[2] Carl Brander et al. “Improving Data-Scarce Image Classification Through Multimodal Synthetic Data Pretraining”. In: IEEE SAS. 2023

[3] Hanna Müller et al. “GAP9Shield: A 150GOPS AI-Capable Ultra-low Power Module for Vision and Ranging Applications on Nano-drones”. In: European Robotics Forum. 2024, pp. 292–297

[4] https://www.bitcraze.io/products/loco-positioning-deck/

[5] https://store.bitcraze.io/products/crazyflie-2-1

[6] Gary Huang et al. “Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments”. In: Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition. 2008

cioflanc@iis.ee.ethz.ch

pulp-platform.org @pulp_platform