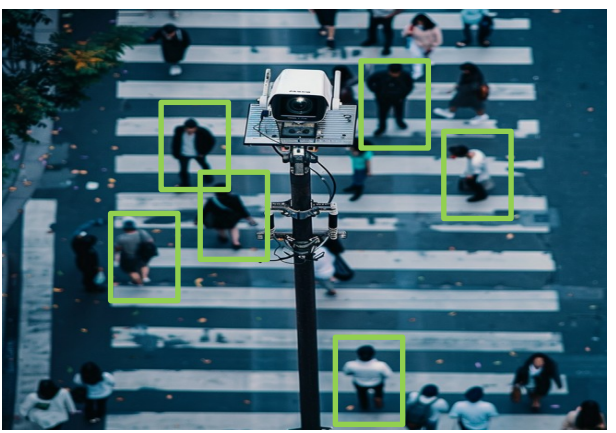# Multi-resolution Rescored ByteTrack for Video Object Detection on Ultra-low-power Embedded Systems

Luca Bompani, Manuele Rusci, Daniele Palossi, Francesco Conti, and Luca Benini
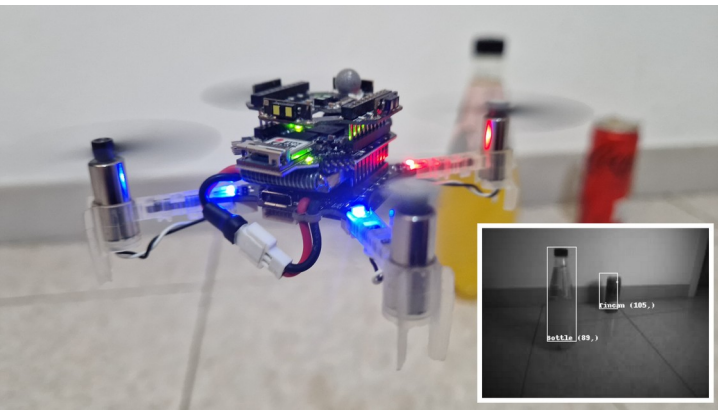
Smart video surveilance

Microcontrollers

State-of-the-art **frame by frame** approaches:

- Tradeoff between frames resolution, detector accuracy and throughput.

- Each frame requires the same amount of computation despite the information carried.

- Each frame is processed indipendently, i.e., no temporal correlation.

Challenge

How to achieve **high-troughtput & high-accuracy video object detection** on **microcontrollers**?

Nano-drones [1]

ETH zürich   ALMA MATER STUDIORUM UNIVERSITÀ DI BOLOGNA

[1] *Lorenzo Lamberti, et al.* Bio-inspired Autonomous Exploration Policies with CNN-based Object Detection on Nano-drones, DATE 2023

Multi-resolution Rescored ByteTrack for Video Object Detection on Ultra-low-power Embedded Systems
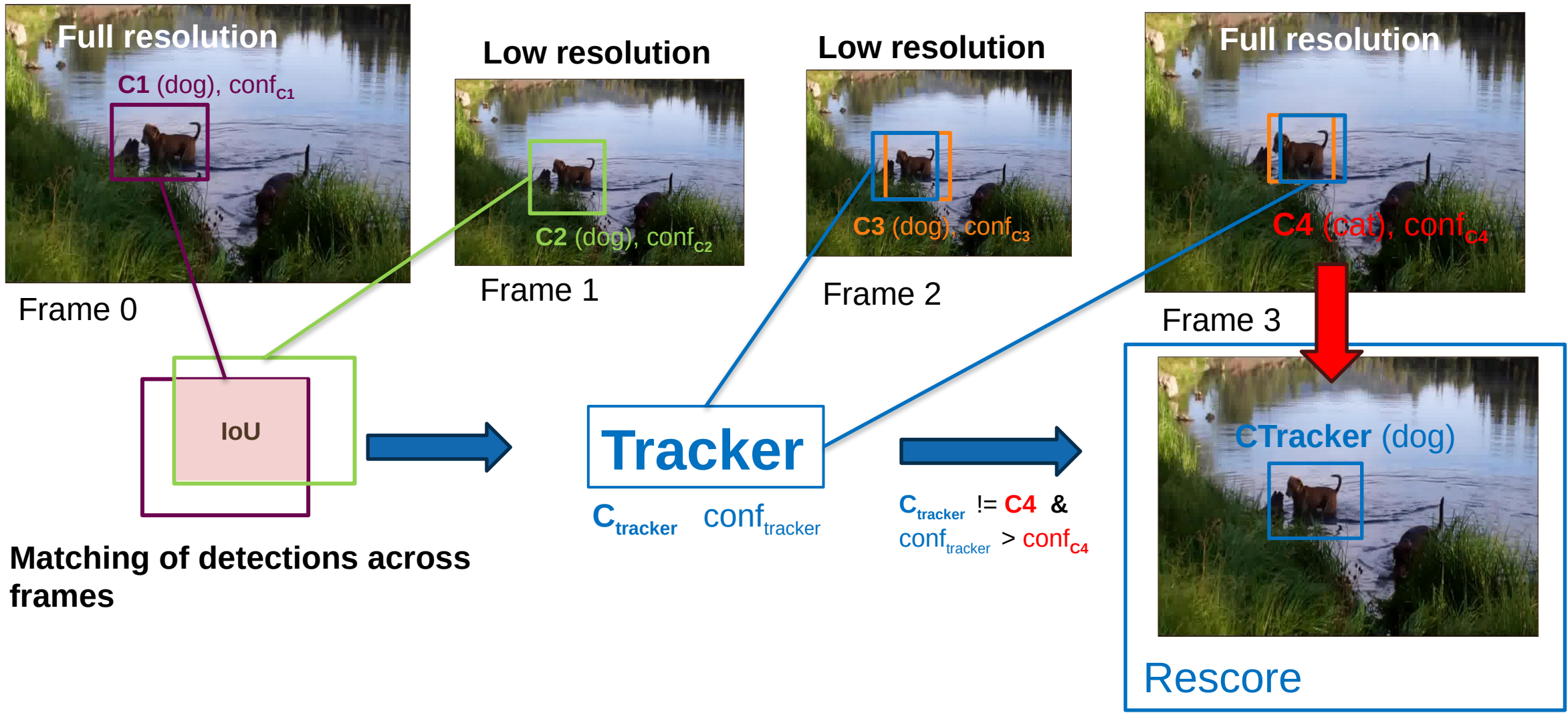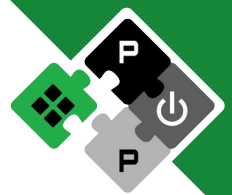
Multi-resolution Rescored ByteTrack for Video Object Detection on Ultra-low-power Embedded Systems

Memory: L1 128 KB, L2 1.5 MB, L3 (off-chip) up to 32 MB.

Cores: 10 cores up to 370MHz, 5 floating point units.

Power: 60 mW, 370MHz (~100x less than a mobile GPU).

SoA Baseline (frame by frame)					MR2-ByteTrack (ours)			
CNN	Input	FPS	mAP	Memory	Input	FPS	mAP	Memory
NanoDet-Plus	320x320	3.3	42.7	2.5 MB	Interleaved 320x320 192x192	5.9	44.9	2.7 MB
	192x192	9.8	27.4	2.5 MB				
YOLOX-Nano	320x320	3.3	41.3	2 MB	Interleaved 320x320 192x192	5.5	41.4	2.1 MB
	192x192	8.6	25.3	2 MB				

Our method, reaches higher accuracy (+2.2 mAP) than frame by frame baseline while increasing troughtput (+2.6 FPS) slightly increasing memory (+7%)

ETH zürich   ALMA MATER STUDIORUM UNIVERSITÀ DI BOLOGNA