

Highlights of architecture optimizations for ultra-low power inference on larger CNNs

Eric Flamand, CTO, GreenWaves Technologies

Some of the architecture choices taken in GreenWaves Technologies next generation GAP IoT Application Processor will be reviewed. This will include a discussion on quantization enhancements, memory bandwidth management and updates to the GAP cluster. It will also look at the evolutions to the GAP SDK and how these exploit the GAP architecture. Finally, some concrete benchmarks on well know NN architectures such as MobileNet V1 and V2 we will review.

The GAP family implements an all in one SoC for sophisticated ML on battery-operated edge devices. It's combination of an MCU-like controller and a flexible but highly optimized compute cluster has set records in its previous generation for AI at the very edge.