



IP library for the acceleration of edge AI/ML

General description

A library with a wide selection of hardware IPs for the design of modular and flexible SoCs that enable end-to-end inference on miniaturized systems. Available IP categories include ML accelerators, dedicated memory systems, the RISC-V based 32-bit processor core icyflex-V, and peripherals.

Our ML accelerators enable parallel computing for dedicated ML tasks, from computer vision to time-series signals classification. The available memory systems are optimized for the accelerators and they are based on either SRAM, register files, or NVM. Thus, the best matching storage solution can be chosen based on the tradeoff among power consumption, memory access, and storage density. A wide range of peripherals enable seamless integration with many external devices.

Tailoring the offered solutions to our customers' needs is our priority at CSEM. Thus, our IPs often allow for design customization and flexible programmability (e.g., for size and precision). The modular nature of this IP library allows for fast and simple integration in any system. A software stack is also available, with firmware examples (e.g., face detection) as well as support for common ML flows and formats (TensorFlow, ONNX, PyTorch, Caffe).

Focus on ML accelerators

- **Binary decision tree (BDT)** for ultra-low-power computer vision, such as presence detection, among other ML tasks. Based on early termination, the BDT minimizes the computational effort based on the image content.
- **Convolutional neural network (CNN)** for complex computer vision tasks. Low-latency and low-power consumption are guaranteed by configurable parallel processing elements (PEs) and programmable precision. Support for several layer types, such as FC, Conv, DSCConv, among others.
- **Recurrent neural network (RNN)** for ML analysis of time-series signals as they capture long-term dependencies. Both LSTM and GRU layer types are supported.

