Object Detection @ 1 mW: Achieving Always-On Computer Vision at the Edge

Ravishankar Sivalingam, Edwin Park, Evgeni Gousev

Qualcomm® AI Research, Qualcomm Technologies, Inc.

Enabling always-on vision, defined by active power of less than 1 mW, is challenging due to power-hungry hardware and the high complexity of inference algorithms. Qualcomm Research has pioneered an Always-on Computer Vision Module (CVM) product based on a holistic approach combining innovations in the system architecture, ultra-low power design, and dedicated hardware for vision algorithms running at the edge. With low end-to-end power consumption, a tiny form factor, and low cost, the CVM can be integrated into a wide range of battery- and line-powered devices (IoT, mobile, VR/AR, automotive, etc.), performing object detection, feature recognition, change/motion detection, and other tasks.

The CVM is built on a custom ASIC, which is a 28nm ultra-low-power ARM-based SoC featuring a control processor, a DSP-like hardware accelerator, a dedicated vision processor, and embedded PMU. It also incorporates a lower-power QVGA CMOS grayscale image sensor and a custom-designed wide field-of-view lens. The image sensor is sensitive to near-IR wavelengths, and can be used for low-light scenarios with IR illumination. The entire CVM, including the image sensor and the ASIC, consumes less than 1 mW power while actively performing computer vision tasks such as object detection. The module is also designed with privacy as a key consideration, since it does not transmit images but rather only the metadata output from computer vision algorithms – e.g., bounding boxes for object detection. A high-level schematic of the CVM is shown in Fig 1.

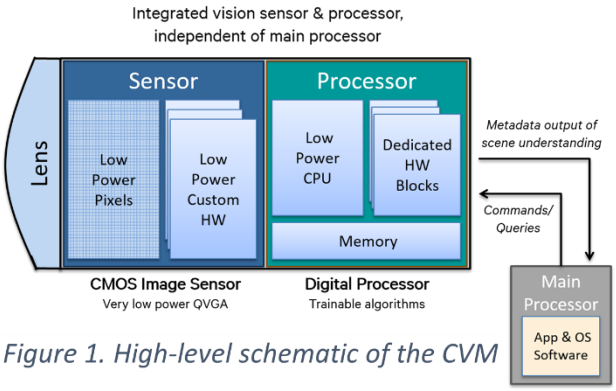

Figure 1. High-level schematic of the CVM

The CVM is capable of running object detection using a scanning-window approach and computer vision features. Dedicated hardware blocks implement efficient feature computation and classification algorithms, enabling always-on object. The DSP-like accelerator is also capable of running compressed neural network models for vision and audio applications. Object detection models for face detection, human detection, certain 3D objects and 2D symbols have been trained and demonstrated for various applications. For example, the module can perform face detection even at extremely low illumination of < 3 lux. The human detection models can detect people as far away as 60 feet. Typical object detection model sizes for such applications range from 10 kB to 100 kB.

As performance requirements may vary depending on the application scenario, in addition to the stock object detection models, we also provide a suite of software tools for product developers so that they may train their own models. These include tools that automatically optimize the runtime performance vs. power of any model, based on the application requirements and scenario-specific data – i.e., for the same model, it allows the user to choose an operating point that has a higher performance accuracy at the cost of more power, or a setting that is much lower power if the use-case allows for slight reduction in accuracy. The optimized model and parameters can then be directly loaded onto the CVM for inference. We will be demonstrating the CVM and software tools at the poster.

Always-on vision as enabled by the CVM will enhance many use-cases across numerous verticals, including smartphones, smart watches, tablets, and AR/VR headsets. IoT applications of the CVM include intelligent occupancy triggers or interactivity triggers for smart home use-cases. The low-cost, low-power, and complete integration of the CVM has enabled new applications that hitherto were not able to use computer vision technology.