

## Pushing the Limits of Ultra-low Power Computer Vision for tinyML Applications

Ravishankar Sivalingam, Edwin Park, Evgeni Gousev

Qualcomm Artificial Intelligence (AI) Research, Qualcomm Technologies, Inc.

Qualcomm's Always-On Computer Vision Sensor (AOCVS) module comprises a low-power monochrome qVGA CMOS image sensor and an ultra-low power custom SoC, with dedicated hardware for computer vision algorithms such as object detection, and is **commercialized as the Qualcomm QCC112 chipset**. The typical end-to-end power consumption (image capture + CV processing, object detection) is only 1 mW.

Through end-to-end optimizations in the data, algorithm, and implementation, we have pushed the limits of object detection performance at ultra-low power. With various enhancements including tuned automatic exposure control, retrained detection models that work well at lower illumination/exposure, and a boosted system clock, we are able to demonstrate **face detection at up to 100 fps** with single-digit mW power consumption. Further, using smart hierarchical chaining of detection models of different object types, we enable the adaptive **detection of multiple objects simultaneously**, without sacrificing frame-rate, while maintaining the same low power profile. This allows us to extend the sensors capabilities to perform simultaneous detection and classification for multiple objects, where the number of objects is only limited by the model memory.

We demonstrate new applications in the **mobile phone/wearables space** and in **IoT markets**, where low-power, low-cost, and high performance are highly desired. Leveraging our rapid and accurate face detection, we enable face-based auto-wake for mobile/wearable screens, user privacy alerts, and trigger other applications such as face authentication. In the **IoT space**, we demonstrate simultaneous shelf status monitoring of products and customer impressions' tracking and analytics. Since we never store or transmit image data, the customers' privacy is always protected.

Users can train object detection models for any object type using our easy-to-use training tools described below and run them on this hardware. Typical model sizes for face detection, human detection, and other applications range between 5 kB to 100 kB. We provide a suite of software tools for users to develop their own custom object detection models, called AOCVS Training Tools. With a **new browser-based GUI front-end, called the AOCVS Portal**, it enables even novice users to rapidly train, tune, and test models for their applications. The user can install this toolset on Windows (natively) or Linux (via Docker containers), either on local machines or cloud-based servers as needed. The web-app interface can be used from any computer.

The software suite includes data preparation tools with a variety of synthetic data augmentation techniques, allowing for training new models with very few positive images of the object. The tools also translate the trained models into a custom format for loading onto hardware. The user is able to optimize the runtime configuration parameters for a given model to customize for their TPR, FPPI, and computational speed requirements. Finally, we also provide a tool for evaluating trained and tuned models on datasets (acquired on the AOCVS module or elsewhere). It supports multiple modes including images, video, and webcam streams, and the user can test the model on pre-existing benchmark datasets, which will be converted to mimic the image quality of the AOCVS module.

We will be demonstrating the above application examples and the AOCVS Portal software at the poster.