Inventec Leverages Qualcomm Dragonwing QCS6490 Processor to Transform Edge AI Applications for Customers

- Compact,fanless edge AI solution: Inventec's AIM-Edge QC01 powered by partner's Hexagon NPU.
- Performance boost : Reduced memory usage by 33% and CPU load by 5%.
- Fast response times : Achieved 18ms per frame for real-time monitoring.
- IoT scalability : Enabled support for multiple camera inputs.
- Improved decision-making : Reduced cloud dependency with enhanced local processing.
- Application sectors : Ideal for safety-critical systems like railway monitoring.

Inventec, in collaboration with Qualcomm Technologies, Inc. has developed the AIM-Edge QC01—a versatile, compact, fanless edge AI box powered by Qualcomm Dragonwing QCS6490 processor. This solution brings cutting-edge AI performance, connectivity, and power efficiency to industrial and commercial IoT applications across various sectors, including smart retail, security, smart transportation, building infrastructure, and smart cities.

One leading global innovation and engineering consulting firm has already leveraged this platform to significantly boost the productivity and efficiency of their railway grade crossing monitoring system. By migrating their AI inference workload to the Inventec AIM-Edge QC01 powered by Qualcomm Hexagon Neural Processing Unit (NPU), the customer reduced memory use by nearly one-third and CPU load by 5%, achieving response times as low as 18 milliseconds per frame. This migration enabled them to reduce cloud dependence, improve local decision-making, and scale to multiple camera inputs, setting new benchmarks for safety-critical railway monitoring.

Introduction

Increasing demand for real-time AI processing at the network edge drives the need for powerful yet energy-efficient, scalable, and flexible edge AI platforms. Dragonwing QCS6490 processor addresses this by combining a Qualcomm Kryo 670 CPU with a Hexagon AI accelerator capable of delivering up to 12.5 TOPS of AI performance under low power consumption, alongside advanced connectivity features like Wi-Fi 6/6E and multi-camera support.

Inventec has harnessed this platform in its AIM-Edge QC01 AI box, designed for fanless operation, compact installation, multi-OS compatibility, and versatile deployment. The device targets a wide range of applications—including queue analytics in retail, slip and fall detection for security, vehicle counting in transportation, cooler door monitoring in

smart buildings, and real-time fleet management—demonstrating the broad utility of this industrial edge AI solution.

Customer’s Railway Crossing Safety Application: A Case Study

One of the most critical safety challenges involves detecting stalled vehicles at railway grade crossings. Timely alerts are paramount to preventing collisions, injuries, and fatalities. The customer developed a sophisticated AI-based video analytics system leveraging the YOLOv8 object detection model, trained via PyTorch and customized for real-time hazard prediction.

Initially, their edge device performed inference locally but still relied heavily on cloud storage and alert transmission, resulting in latency, network traffic, and memory footprint concerns. Moreover, the capacity to process concurrent video streams from multiple cameras was limited, constraining deployment in complex railway environments.

To overcome these limitations, the customer migrated their solution onto Inventec’s AIM-Edge QC01 platform, powered by the Dragonwing QCS6490 processor. This integration delivered:

- Substantial Memory and CPU Utilization Reductions: Memory usage decreased by 32.92%, CPU load dropped by 5%, while inference performance reached 18 ms per frame.
- Multi-Camera Scalability: The triple image signal processor (ISP) on the Dragonwing QCS6490 supports up to five concurrent 4K video streams, enabling the system to handle complex rail safety scenarios across multiple vantage points.
- Reduced Cloud Dependence: Edge AI inference reduces network traffic and communication latency, lowering overall solution costs by approximately 30%.

- Efficient Neural Processing: The Hexagon NPU handles AI workloads, freeing CPU and GPU resources for other system functions.
- Long Lifecycle Support: Enterprise-grade hardware guarantees prolonged OS and security updates, ensuring robustness for industrial deployments.

Technical Integration Highlights

The customer efficiently adapted their AI model pipeline to the Qualcomm Technologies platform without altering the base model. Key steps included:

- Model Conversion: Changing the YOLOv8 model from ONNX to Qualcomm Technologies' DLC format, with quantization fine-tuned to balance accuracy (FP16 activations, INT8 weights).
- Graph Caching: Using Qualcomm Neural Processing SDK tools to generate offline graph caches, reducing AI model initialization latency.

- Deployment: Running the Snapdragon Neural Processing Engine enabled software on the AIM-Edge QC01 under Ubuntu, interfacing via Snapdragon Neural Processing Engine Helper (a Python API wrapper) for inference execution.

This streamlined approach allows customer to integrate additional railway safety models (crowd monitoring, weapon and violence detection) rapidly, accelerating deployment timelines.

Inventec AIM-Edge QC01 Platform Overview

Feature	Description
SoC	Dragonwing QCS6490 (Qualcomm Kryo 670 CPU + Hexagon AI)
AI Performance	Up to 12.5 TOPS (NPU)
Connectivity	Enterprise-grade Wi-Fi 6/6E, Ethernet
Camera Support	Up to 5 concurrent inputs at 4K30/4K60 resolution
Form Factor	Compact, fanless, lightweight
Power Efficiency	Low power consumption, long lifecycle for industrial use
Software Support	Qualcomm Neural Processing SDK, Qualcomm AI Runtime (AIRT)

Key Applications	Industrial IoT, smart retail, security, transportation, smart buildings, railway safety

Industry Impact and Future Outlook

Through the synergy of Inventec’s deployment expertise and Qualcomm Technologies’ heterogeneous computing leadership, the AIM-Edge QC01 enables scalable, real-time AI applications that reduce latency, enhance privacy, and lower operational costs compared to cloud-heavy approaches.

The customer’s rail safety success exemplifies the platform’s capability to deliver mission-critical AI inference with low latency and high efficiency, while supporting multi-camera and multi-model scenarios.

Both Inventec and Qualcomm Technologies continue to expand this ecosystem with developer tools, software integrations, and additional AI models—paving the way for democratized edge AI adoption across industries, making intelligent and safe edge environments a reality worldwide.