CNN-Based QR Code Feature Extractor: Hardware Accelerator Project Report

Overview:

This project aims to accelerate QR code feature extraction using a CNN-based approach, where convolutional layers and dense operations are offloaded to a custom hardware chiplet implemented on FPGA. The goal is to reduce latency and improve real-time performance on edge devices.

Heilmeier Questions Summary:

- What are you trying to do?

  Design and implement a CNN-based system that extracts features from QR codes using hardware acceleration.

- How is it done today and what are the limits?

  Currently handled via CPU/GPU inference which limits deployment on power-constrained or real-time embedded devices.

- What is new in your approach?

  We implement a lightweight CNN and move MAC operations to a custom Verilog-based chiplet on FPGA.

- Who cares?

  Real-time AI use-cases in logistics, AR, drone scanning, and embedded security systems.

- What are the risks?

  FPGA resource limits, integration complexity, and potential communication overhead.

- How much will it cost?

  $0 using department resources and open-source tools.

- How long will it take?

  10-11 weeks, spanning model development, profiling, hardware implementation, and integration.

- What are the midterm/final exams?

  Midterm: Working Python CNN, profiling results. Final: Hardware-accelerated inference working on FPGA

with measurable speedup.

Analysis:

Profiling with cProfile shows that >80% of the compute time is consumed in convolution and ReLU layers. These are highly regular and parallelizable, making them ideal for hardware acceleration.

A NetworkX-generated dataflow graph visualizes dependencies and confirms modular computation structure. Doxygen's call graph provides function linkage clarity.

System Block Diagram:

[Python] --> [Preprocess] --> [FPGA: CNN Accelerator] --> [Feature Output]

Future Work:

- Quantization for smaller footprint

- AXI-based high-speed interface

- Expand to QR decoding, not just feature extraction

Conclusion:

The project establishes a hardware/software co-design pipeline that significantly accelerates CNN-based QR analysis. This proves useful for scalable and efficient edge AI.