Michal Danilowicz, Tomasz Kryjak

This document provides a detailed academic summary of the research paper "Real-Time Multi-Object Tracking using YOLOv8 and SORT on a SoC FPGA" by Michal Danilowicz and Tomasz Kryjak. The analysis delves into the paper's contributions, methodology, findings, theoretical underpinnings, limitations, future research directions, interdisciplinary implications, and a conclusive assessment. Each section aims to provide a comprehensive and intellectually rigorous exploration of the presented work.

The research paper makes several significant contributions to the field of embedded computer vision and real-time multi-object tracking (MOT). The core contribution lies in the successful implementation of a high-performance MOT system on a System-on-Chip Field-Programmable Gate Array (SoC FPGA), specifically the Zynq UltraScale+ MPSoC from AMD/Xilinx. This achievement addresses a critical need for efficient and low-power MOT solutions for applications like autonomous vehicles, advanced driver-assistance systems (ADAS), and advanced video surveillance systems (AVSS), where real-time performance and energy efficiency are paramount.

The authors' approach distinguishes itself from prior work in several key aspects:

- **High-Quality Object Detection:** Utilizes the state-of-the-art YOLOv8 object detection model, balancing accuracy with resource constraints using YOLOv8n (nano).
- **Quantization-Aware Training (QAT):** Employs QAT using the Brevitas library to optimize YOLOv8 for efficient integer-only inference on the FPGA, mitigating accuracy loss.
- **Hardware-Software Co-design:** Offloads computationally intensive object detection to the FPGA's programmable logic (PL), while the tracking algorithm (SORT) runs on the processor system (PS).
- **FINN Framework Adaptation:** Adapts the FINN framework to accommodate the YOLOv8 architecture and efficiently manage external memory for model parameters, including modifications for channel split and concatenation.
- **Comprehensive Evaluation:** Evaluates the detector's performance using the COCO dataset (mAP metric) and the overall MOT system's performance using the MOT15 dataset (MOTA and CLEARMOT metrics), using private detections.

In summary, the paper presents a novel and highly effective approach to real-time embedded MOT, combining a state-of-the-art object detector with a robust tracking algorithm, optimized for efficient hardware implementation on an SoC FPGA. The adaptation of the FINN framework and the comprehensive evaluation methodology further strengthen the paper's contributions.

The methodological architecture is characterized by a multi-faceted approach encompassing model selection, optimization, hardware implementation, and rigorous evaluation. The authors meticulously detail each stage, providing a clear and reproducible framework for future research.

Key methodological elements include:

- **Model Selection:** Strategic choice of YOLOv8n for its balance between accuracy and computational efficiency.
- **Model Optimization (Quantization-Aware Training):** Implementation of QAT using the Brevitas library within PyTorch, with 4-bit quantization of weights and activations and EMA during fine-tuning.
- **Hardware Implementation (SoC FPGA):** Leveraging the heterogeneous architecture of the Zynq UltraScale+ MPSoC, offloading object detection to the PL and running SORT on the PS, utilizing DMA modules for efficient data transfer.
- **FINN Framework Adaptation:** Addressing challenges of integrating YOLOv8's architecture into FINN, modifying the framework to handle specific operations and data flow, and optimizing operations within FINN.
- **Evaluation Methodology:** Comprehensive evaluation using COCO (mAP) and MOT15 (MOTA and CLEARMOT) datasets, with private detections for realistic assessment.

The methodological rigor, combined with the innovative use of existing tools and frameworks, makes this research highly valuable and reproducible.

The research yields a hierarchy of critical findings, demonstrating the effectiveness and efficiency of the proposed embedded MOT system.

- **High-Speed Object Detection:** Achieved a processing speed of 195.3 fps on the ZCU102 SoC FPGA at 300 MHz.
- **Acceptable Detection Accuracy:** Maintained an mAP of 0.21 on the COCO validation set despite 4-bit quantization.
- **Robust Multi-Object Tracking:** Achieved a MOTA of 0.389 on the MOT15 dataset, exceeding the original SORT algorithm with Faster-RCNN.
- **Comparable Performance to State-of-the-Art (on a subset):** Exhibited comparable MOTA scores to [21] on a subset of ua-detrac.
- **Resource Utilization:** Reported FPGA resource utilization (74.71% LUTs, 44.94% FFs, 46.16% BRAMs, 19.29% DSPs).

These findings collectively demonstrate the success of the proposed approach in achieving real-time, high-performance MOT on an SoC FPGA.

The research integrates several theoretical frameworks from computer vision, deep learning, and embedded systems design.

- **Tracking-by-Detection Paradigm:** Validates the effectiveness of this paradigm in embedded systems.
- **Deep Learning for Object Detection:** Supports the effectiveness of deep learning for object detection in resource-constrained environments.
- **Quantization Theory:** Demonstrates the potential of quantization techniques to enable deep learning on embedded systems.
- **Kalman Filtering and Data Association:** Validates the effectiveness of Kalman filtering and the Hungarian algorithm in real-time MOT.
- **Hardware-Software Co-design Principles:** Demonstrates the effectiveness of this approach in achieving real-time performance.
- **FINN Framework and High-Level Synthesis (HLS):** Showcases the power of HLS in accelerating deep learning inference on FPGAs.

The successful integration of these theoretical frameworks demonstrates the paper's contribution to the advancement of embedded computer vision and real-time MOT.

The research has several limitations:

- **Dataset Limitations:** Evaluation primarily based on COCO and MOT15 datasets, limiting generalizability.
- **Quantization Effects:** 4-bit quantization introduces a noticeable drop in detection accuracy.
- **Algorithm Simplicity:** Use of SORT limits the system's ability to handle complex tracking scenarios.
- **Hardware Platform Specificity:** Implementation is specific to the Zynq UltraScale+ MPSoC platform.
- **Software Optimization:** Python implementation of SORT and NMS on the PS may not be the most efficient.
- **Limited Evaluation Metrics:** Could include additional metrics for a more nuanced understanding of performance.

These limitations highlight the epistemological boundaries of the research. The findings are valid within the specific context of the chosen models, datasets, and hardware platform.

Several promising avenues for future research emerge:

- **Advanced Tracking Algorithms:** Exploring more sophisticated tracking algorithms (DeepSORT, ByteTrack).
- **Improved Quantization Techniques:** Investigating advanced quantization techniques (mixed-precision, learned quantization).
- **Hardware Architecture Optimization:** Optimizing the hardware architecture (systolic arrays).
- **Platform Portability:** Developing a more platform-independent implementation.

- **Software Optimization (PS):** Optimizing the software implementation using C or C++.
- **Real-time Video Streaming Integration:** Integrating with a real-time video streaming pipeline.
- **Power Consumption Analysis:** Detailed analysis of power consumption.
- **Robustness to Adverse Conditions:** Testing robustness to varying lighting conditions, occlusions, and camera motion.

These research directions offer significant opportunities to build upon the foundation laid by this paper.

The research has significant implications across multiple disciplines:

- **Computer Vision:** Advances embedded computer vision, enabling deployment of sophisticated algorithms on resource-constrained devices.
- **Deep Learning:** Demonstrates the feasibility of deploying quantized deep learning models on embedded hardware.
- **Embedded Systems:** Showcases the importance of hardware-software co-design in developing efficient embedded systems.
- **Robotics:** Development of efficient and low-power MOT systems is crucial for robotics applications.
- **Autonomous Driving:** Direct implications for autonomous driving, where real-time object tracking is essential.
- **Video Surveillance:** Suitable for battery-powered surveillance cameras.
- **Hardware Acceleration:** Demonstrates the effectiveness of FPGA-based hardware acceleration for deep learning inference.

The interdisciplinary nature of this research highlights the importance of collaboration between different research communities.

The research paper "Real-Time Multi-Object Tracking using YOLOv8 and SORT on a SoC FPGA" presents a significant contribution to the field of embedded computer vision and real-time MOT. The authors successfully demonstrate the feasibility and effectiveness of deploying a high-performance MOT system on a resource-constrained SoC FPGA platform. The combination of a state-of-the-art object detector (YOLOv8n), a robust tracking algorithm (SORT), and an efficient hardware implementation using the adapted FINN framework results in a system that achieves real-time performance with acceptable accuracy.

The comprehensive evaluation methodology provides a more realistic and robust assessment of the system's capabilities. The findings validate the effectiveness of the tracking-by-detection paradigm, QAT, and hardware-software co-design principles in the context of embedded MOT. While the research has limitations, these limitations provide valuable directions for future research. The high processing speed of the detector, the acceptable accuracy despite quantization, and the robust tracking performance collectively demonstrate the potential of the proposed approach for a wide range of real-world applications. The paper's detailed methodology and comprehensive evaluation make it a valuable contribution to the field.