Development of an AI-Based Visual Perception System with FPGA Acceleration for Object Detection and Tracking on UAVs

Program Riset ITB 2026 dan Skema Riset Dosen Muda ITB DRI.PN-6-79-2026

Anggera Bayuwindra, Ph.D[1]., Nana Sutisna, Ph.D. [1], Hilton Tnunay, Ph.D. [1], Dr. Sahat Hutajulu[2],
Galang Adira Prayoga M.T[1], Iqbal Najmul Achyar, S.T[1], Miqdad Abdurrahman, S.T[1].
[1]Sekolah Teknik Elektro dan Informatika (STEI), [2]Sekolah Bisnis dan Manajemen (SBM)

Abstract

The rapid development of UAV technology demands fast, efficient, and accurate autonomous visual perception systems. To overcome the severe computational, latency, and energy constraints of running deep learning models onboard, this research proposes the development of an Artificial Intelligence (AI) visual perception system accelerated by FPGA. Utilizing the UAV primarily as an experimental platform, this study methodologically evaluates the co-design between hardware-aware AI algorithms and FPGA architectures. The methodology encompasses AI optimization via model quantization and compression, alongside a rigorous trade-off analysis between accuracy and computational efficiency through simulations and limited flight tests. Expected outcomes include an experimentally validated AI-FPGA framework and international publications, ultimately establishing a robust technical foundation for the future development of embedded AI and autonomous UAV systems.

Keyword: *Unmanned Aerial Vehicle (UAV); Artificial Intelligence; Machine Learning; FPGA Acceleration; Vision-Based Target Detection; Hardware-Aware AI; Embedded Systems*

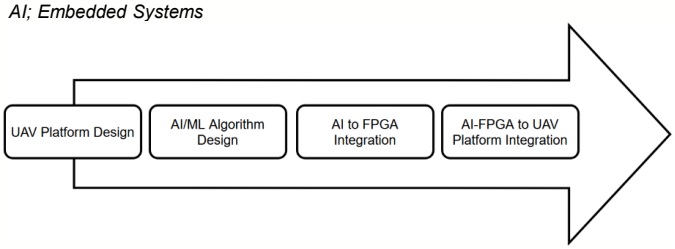


Fig 2. Outlines of the systematic research methodology workflow, progressing through four main stages: UAV platform design, AI/ML algorithm design, AI to FPGA integration, and culminating in the final AI-FPGA to UAV platform integration.



Fig 1. Various practical use cases of Unmanned Aerial Vehicles (UAVs) in dynamic environments, specifically highlighting applications in border surveillance, critical infrastructure monitoring, environmental surveys, and disaster response operations.



Fig 6. Comparison of tank detection results using RGB and thermal imagery, showing the original images (left), RGB images with YOLO bounding box detections (middle), and thermal images with corresponding bounding box detections (right).

Research Method

This research integrates AI algorithms, FPGA hardware acceleration, and a UAV testbed to evaluate vision-based target detection and localization. The methodology uses the UAV purely as an experimental platform to first baseline AI models, then applies hardware-software co-design optimizations like quantization for real-time FPGA processing. Finally, the integrated system is validated through simulations and limited flight tests to assess accuracy, latency, and efficiency trade-offs, yielding a TRL 2-3 proof-of-concept for future embedded AI systems.

Discussion & Result

The current state of this research has successfully validated the integration of the designed AI/ML algorithms with a GPU-based onboard computational unit. During active flight operations, the UAV's visual sensors stream continuous image data directly to this embedded system. The compute unit then processes this incoming feed to perform real-time object detection and classification within a dynamic aerial environment.



Fig 3. Offline tracking results generated by the AI model, showcasing successful object detection and bounding box placement on multiple targets using a top-down vehicle dataset.



Fig 4. Live online tracking capabilities of the visual perception system, illustrating the system successfully detecting and tracking a pedestrian in real-time during active flight.

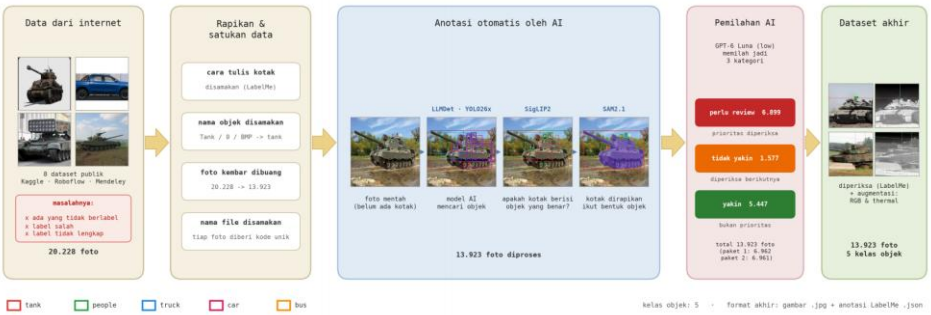


Fig 5. Overview of the dataset preparation and annotation pipeline, including public image collection, data cleaning and standardization, AI-assisted annotation using multiple models, AI model selection based on performance evaluation, and final dataset preparation with LabelMe annotations for five object classes.

This successful GPU-based implementation serves as a crucial baseline, demonstrating the technical viability and accuracy of the autonomous visual perception system. By establishing this functional foundation, the research provides a clear comparative benchmark for subsequent phases, particularly for evaluating the latency and energy efficiency improvements when transitioning to highly optimized, FPGA-accelerated architectures.

Conclusion

This research demonstrates the feasibility of an AI-based visual perception system for UAV applications, achieving real-time object detection, classification, and tracking using onboard visual sensors and GPU-based processing. The developed system has been validated through offline evaluations and live flight experiments, establishing a functional baseline for autonomous visual perception. Furthermore, the systematic dataset preparation and AI-assisted annotation pipeline provide a foundation for developing and evaluating robust detection algorithms under diverse operating conditions.

Building upon this validated baseline, the next phase focuses on hardware–software co-design to integrate optimized AI algorithms with FPGA architectures. Techniques such as model quantization and compression will be investigated to reduce computational complexity, latency, and energy consumption while maintaining acceptable detection accuracy. This development will establish the foundation for efficient FPGA-accelerated embedded AI systems, supporting future advances in autonomous UAV perception and multi-modal sensor integration.