Dynamic Power Management for Edge AI: A Sustainable Self-Adaptive Approach

The rapidly growing deployment of Edge AI devices performing high-demand tasks, such as real-time object detection, creates a critical challenge: balancing high performance (maintaining a target confidence) against the severe constraints of intermittent power supply from solar energy harvesting. This thesis addresses the necessity for a dynamic policy that can effectively manage this dual-objective trade-off over long operational horizons. The research establishes an empirical foundation via a parameter study conducted on Raspberry Pi hardware, quantifying the stochastic relationship between configuration parameters (model variant, resolution, frame rate) and actual power consumption/detection confidence, which revealed median shifts of up to 1.37W in power consumption and up to 26 percentage points in detection confidence between different operational configurations. This data informed the construction of a custom Reinforcement Learning (RL) environment that utilizes Kernel Density Estimation (KDE) to model hardware stochastically and physics-based models for solar dynamics. To solve the dual-objective problem of maximizing performance while satisfying the long-term survival goals, a Proximal Policy Optimization (PPO) agent was trained within a Constrained Optimization framework. The agent's policy was rigorously evaluated over 24-hour and 48-hour cycles across six dynamic scenarios against static and random baselines. The results confirm that the PPO agent successfully learned an adaptive strategy: it consistently manages the trade-off better than non-learning baselines, strategically scaling its resource use based on real-time energy context. Quantitative analysis showed that the PPO agent survived up to 1.5 hours longer than the more power-hungry baselines while achieving at least 40 percentage points more SLA satisfaction than the least power-hungry static policy. This work provides a validated, data-driven approach for sustainable resource management in energy-constrained Edge AI systems.