Metareasoning Approaches to Thermal Management During Image Processing

Michael K. Dawson Jr., Jeffrey W. Herrmann

Department of Mechanical Engineering, A. James Clark School of Engineering, University of Maryland, College Park

MARYLAND ROBOTICS CENTER

THE INSTITUTE FOR SYSTEMS RESEARCH

Introduction

Resource-constrained electronic systems are present in many semi- and fully-autonomous systems and are tasked with computationally heavy tasks such as image processing. Without sufficient cooling, these tasks often increase device temperature up to a predetermined maximum, beyond which the task is slowed by the device firmware to maintain the maximum. This is done to avoid decreased processor lifespan due to thermal fatigue or catastrophic processor failure due to thermal overstress. The algorithms that control this thermal throttling are implemented at a firmware level and not consistent across devices.

In image processing, there is a trade-off between processing speed and detection precision of neural network object detectors.

Metareasoning is a method for altering an agent's decision-making process by changing parameters or switching between decision-making algorithms to increase the overall performance of the agent.

This research attempted to leverage the precision-throughput tradeoff of object detectors to maintain a stable temperature using metareasoning.

Approach

The following three algorithms were created and executed in Python3:

- Throughput Adjustment Policy (1)
- Network Switching Policy (2)
- Hybrid Policy (3)

Networks EfficientDet D0 - D4, optimized for mobile platforms via TensorFlow Lite, were used for object detection. They were trained on the COCO image dataset.

Metareasoning occurred after image processing and could add pauses and switching between networks. It had knowledge of the temperature, pause length history, network usage history, average expected detection precision, and previous pause lengths.

Tests were performed on the following devices:

- Raspberry Pi 4B
- NVIDIA Jetson Nano Developer Kit

Results

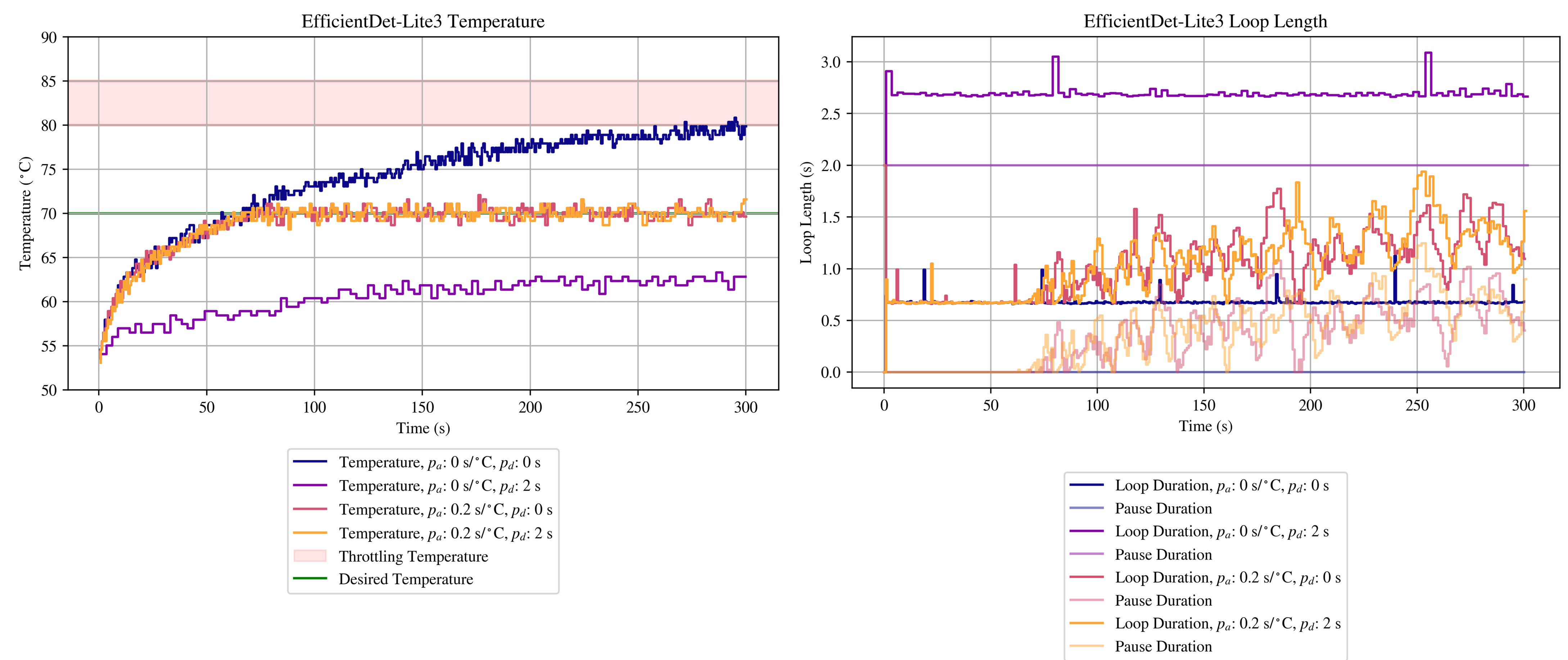

Figure 1: Throughput Adjustment Policy (a) temperature and (b) throughput for selected networks.

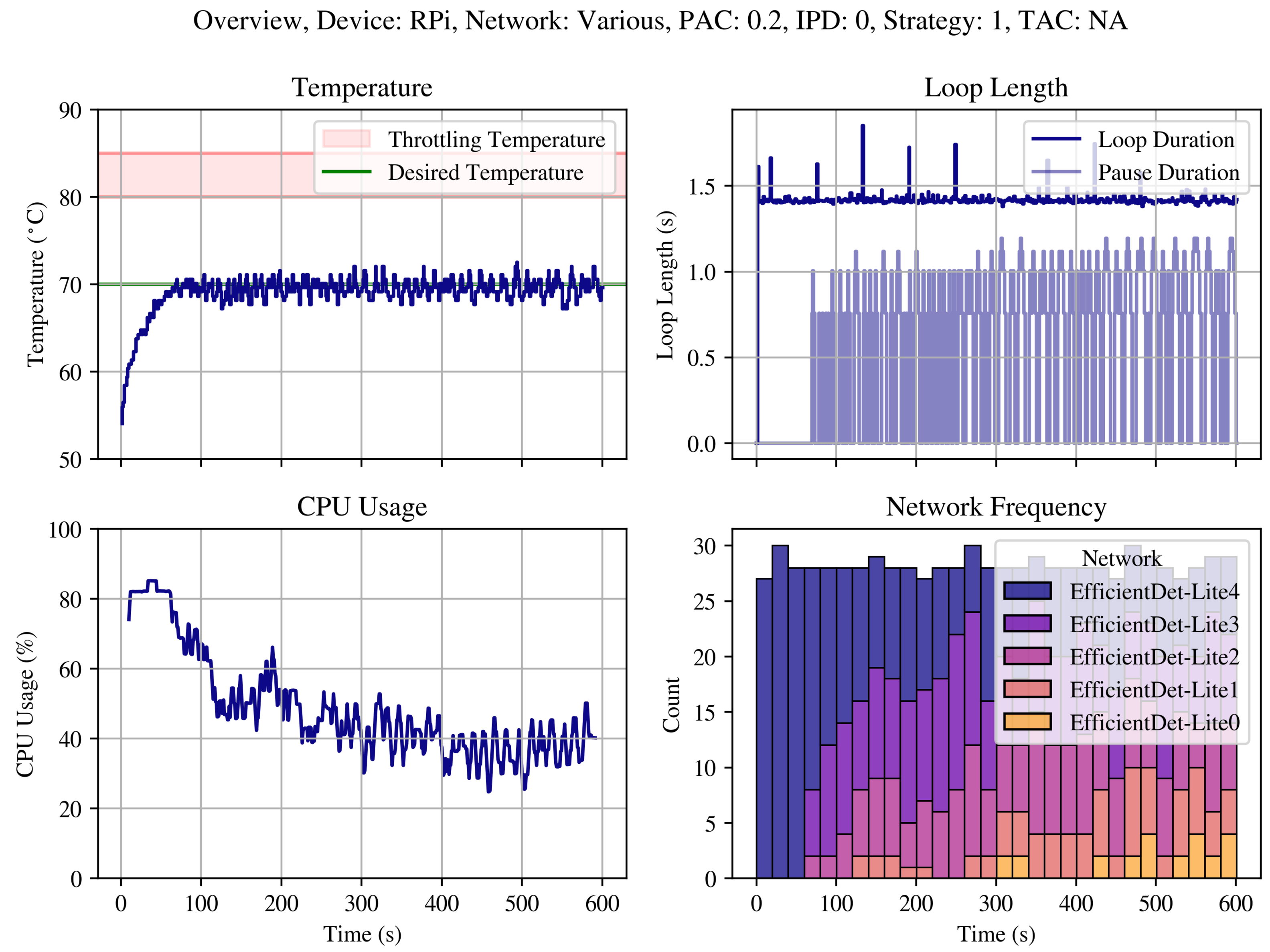

Figure 2: Network Switching Policy metrics.

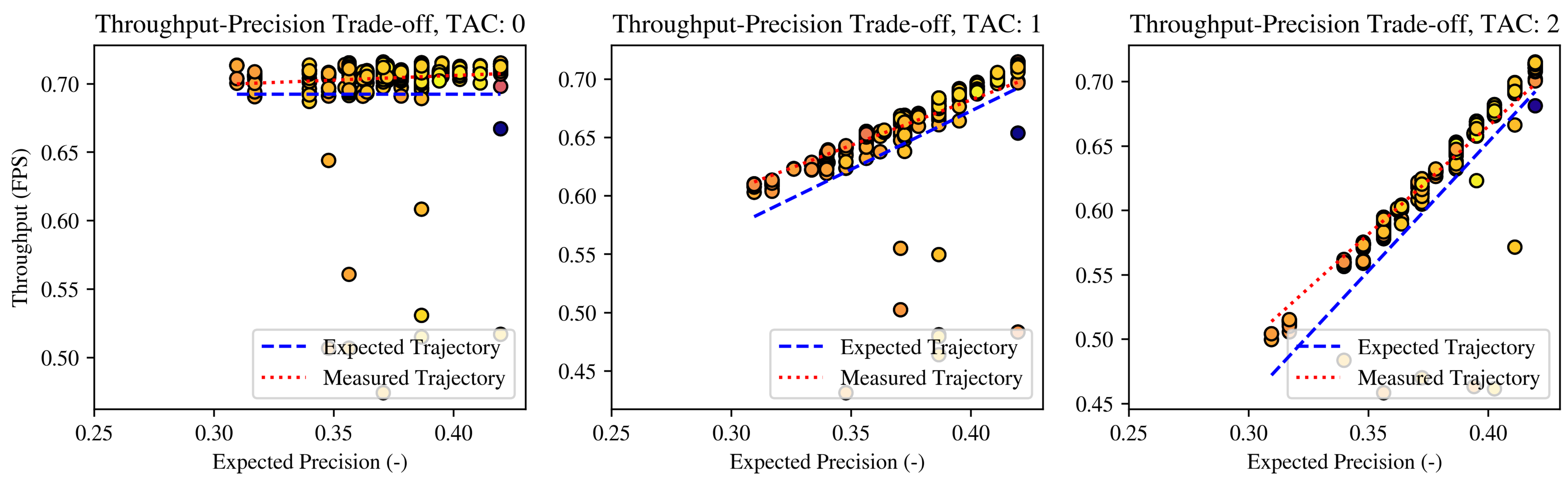

Figure 3: Hybrid Policy trade-off demonstration.

Conclusions

Testing results show that policies 1 and 2 are successful at maximizing either detection precision or minimizing throughput variance for a range of parameter values. Results for policy 3 show that the throughput adjustment coefficient (TAC) parameter can be tuned to maintain a constant temperature while performing a user-defined trade-off between detection throughput and precision.

These results demonstrate that metareasoning can provide variable precision or throughput for thermal throttling with only a small loss in overall efficiency.

Additional Resources

Dawson, *Metareasoning Approaches to Thermal Management During Image Processing.* 2022

Bianco et al., *Benchmark analysis of representative deep neural network architectures.* 2018

Huang et al., *Speed/accuracy trade-offs for modern convolutional object detectors.* 2017

Lee et al., *Benchmarking video object detection systems on embedded devices under resource contention.* 2017

Lee et al., *Virtuoso: Video-based intelligence for real-time tuning on socs.* 2021

Tan et al., *Efficientdet: Scala bale and efficient object detection.* 2019

Abadi et al., *TensorFlow: Large-scale machine learning on heterogenous systems.* 2015

Cox and Raja, *Metareasoning: Thinking About Thinking.* 2011

COCO Consortium, *COCO – Common Objects in Context.* 2017

Acknowledgments

The NVIDIA Jetson Nano Developer Kit was provided by the UMD Robotics Realization Laboratory.

This research is supported by U.S. Army Grant No. W911NF2120076. We appreciate the guidance and feedback of our ARL collaborators.

2nd Annual Maryland Robotics Center Research Symposium, May 31, 2022