Real-Time Object Detection on Low-Power Edge Device with Hardware Acceleration

Graduate



Marc Benz

Introduction: Convolutional Neural Networks (CNNs) have led to significant advancements in the field of computer vision, including object detection. However, the computational complexity of CNNs requires significant electrical energy per inference cycle when executed on standard GPUs. This and the significant cost of GPUs prevent the application of CNNs in many practical applications. A resolution to this issue is aimed by specialized hardware that is at the same time power efficient, fast and low cost. This thesis explores the use of a dedicated accelerator in combination with a microcontroller on an embedded edge device for low-power object detection.

Approach: The target platform selected is the Google Coral Dev Board Micro. The board is equipped with an Edge TPU, a specialized component designed for accelerated AI processing designed by Google. It includes a microcontroller as its main processor and a camera for image acquisition. With the help of the Keras framework, a CNN was implemented and trained on a PC. The weights were then quantized to 8-bit integers and compiled for the Edge TPU.

Conclusion: The capabilities of the Edge TPU on the Google Coral Dev Board Micro were evaluated on a practical task. The results show that this board is capable of running the inference in real-time with a latency of around 20ms, while consuming less than 1W of power.

Google Coral EdgeTPU compile workflow. https://coral.ai/docs/edgetpu/models-intro/ Application: Prediction map of a cyclist detection model. Own presentment



The detection model contains operations not supported by the TPU, these operations are performed by the CPU. https://coral.ai/docs/edgetpu/models-intro/





Advisor Prof. Dr. Martin Weisenhorn

Co-Examiner

Prof. Dr. Helmut Grabner, ZHAW Zürcher Hochschule für angewandte Wissenschaften, Winterthur, ZH

Subject Area

Data Science, Computer Science, Electrical Engineering

