

On-Site Scene Understanding with Edge AI

Student



Lukas Schöpf

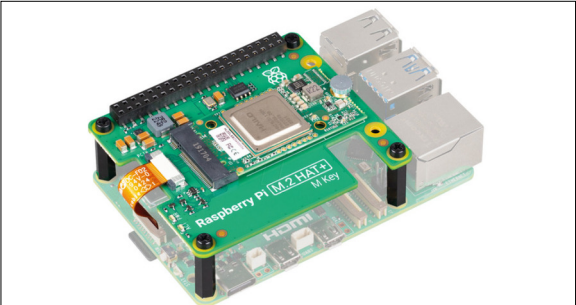
Introduction: Hexagon AB, a leader in digitization technology, focuses on integrating advanced scene understanding capabilities into edge devices for its scanning applications. This project evaluated the use of hardware accelerators, identifying Hailo's products as the most suitable due to their availability, power efficiency, and computational performance. By using Hailo's M.2-enabled hardware, neural networks can be efficiently run on Raspberry Pi.

Approach: The study builds on previous work by Lia Winkler, who demonstrated that CLIP, a model for aligning text and images, is well-suited for scene classification tasks. Her best practice involved dividing panoramic images into patches and classifying them via majority voting, achieving strong results. TinyCLIP, a lightweight version of CLIP, was also mentioned for its reduced parameter count and efficiency. The core of this project centered around implementing the CLIP and TinyCLIP models on Hailo hardware to enable efficient scene understanding. CLIP, used to align text prompts with images, demonstrated strong performance in prior research. Due to limitations in the Hailo platform's ability to compile transformer-based architectures, pre-calculated text embeddings were used, and ResNets, a special CNN architecture, employed as vision encoders. Challenges arose in maintaining model accuracy after quantization, particularly for models with a lower parameter count. Various strategies were attempted to mitigate accuracy loss, such as adjusting quantization levels, reworking the network split, and fine-tuning classification thresholds. However, only the adjusted thresholds showed a slight improvement in some models.

Result: The results indicated that model accuracy after quantization is highly dependent on the

network's parameter count. Interestingly, TinyCLIP-30M, with fewer parameters, showed better post-quantization accuracy than the larger RN50 model, suggesting that TinyCLIP's optimization effectively identifies the most important model weights.

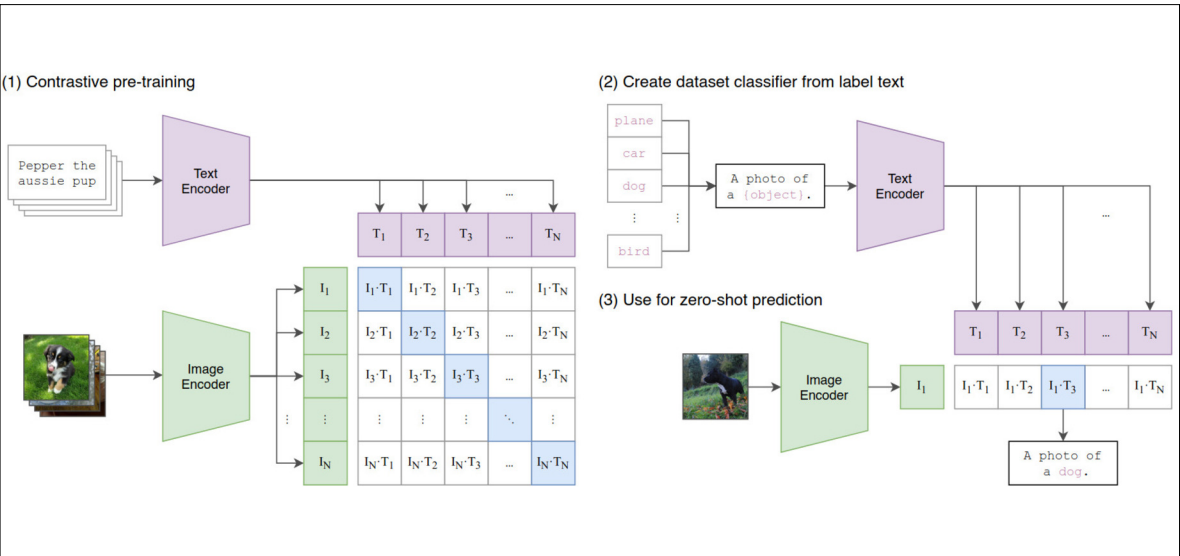
Raspberry Pi with M.2 HAT+ equipped with the Hailo-8L hardware accelerator.
<https://www.pi-shop.ch/raspberry-pi-ai-kit>



Model performance for new cut of the model graph. The impact of the improved threshold can be seen in the last column.
Own presentation

Modelname	Accuracy	Balanced accuracy	Change in balanced accuracy
RN50	0.799	0.200	0
RN50x4	0.845	0.560	0
RN101	0.743	0.507	+0.93
TinyCLIP-19M	0.141	0.211	-0.06
TinyCLIP-30M	0.624	0.521	+0.95

Overview of how CLIP (Contrastive Language-Image Pre-Training) works.
<https://arxiv.org/abs/2103.00020>



Advisor

Prof. Dr. Andreas Breitenmoser

Subject Area

Electrical Engineering,
Data Science

Project Partner

Hexagon AB