Embedded machine learning for microcontrollers

09 January 2024, 15:44
Frank Van den Broek

Embedded machine learning put to practice: computer vision on a Raspberry Pi Pico microcontroller

Until quite recently, the domain of edge machine learning (edge ML) mostly focused on inference either on actual computers with GPUs or on mobile devices. With the rise of tiny machine learning (TinyML), the deployment of neural networks and data analytics will also be possible on microcontrollers. To demonstrate the possibilities of TinyML, Sirris has rolled out a computer vision application on a low-cost microcontroller: the Raspberry Pi Pico.

TinyML is a fast-growing multidisciplinary technology where innovations in embedded hardware, software and machine learning enable a new class of smart applications through on-device, near-sensor inference on ultra-low power embedded systems. This opens the door to new types of edge services and applications that do not depend on cloud processing, but are based on inference performed on the edge. Because data is processed on-device, and not sent over a network, TinyML offers great advantages in terms of privacy, latency, energy efficiency and reliability. A well-known first use case of TinyML on always-on, battery-powered devices was the so-called audio wake word detection on products from tech giants ("Okay Google", "Hey Siri"). Thanks to recent evolutions in TinyML development platforms, machine learning models and embedded hardware, embedded machine learning is now also within reach of the broader industry for many different

applications that involve sensors such as cameras, accelerometers and sound sensors.
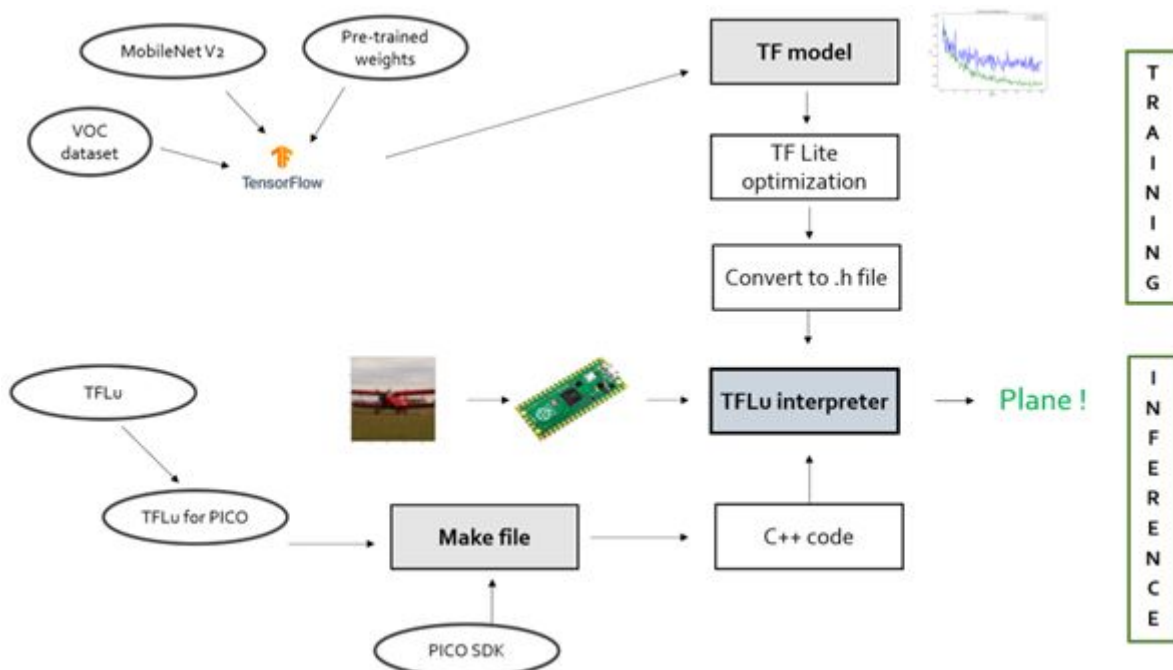
Computer vision on a low-cost microcontroller

To demonstrate the possibilities of TinyML, Sirris has rolled out a computer vision application on a low-cost microcontroller: the Raspberry Pi Pico (costs less than 5 euros). The aim was to assess the achievable performance, such as accuracy and inference time for an image classification task, using open source computer vision models and tools.

Computer vision has seen a major breakthrough in the past decade thanks to deep learning algorithms, AI networks that used to require a lot of memory and computing power and therefore could not be applied on microcontrollers. For example, the Raspberry Pi Pico only features 264 KB SRAM / 2 MB flash memory and a 133 MHz Arm Cortex-M0+ processor. Now, specifically for mobile and embedded devices, new deep learning network architectures have been developed that perform well for computer vision tasks, even with limited resources. For our case, we used a MobileNetV2, a pre-trained convolutional neural network that is freely available and that we further trained on the VOC-2012 dataset for our application via transfer learning.

Open source tools

Training a TinyML model is done with a similar training process as for larger deep learning models. For our application, we used Tensorflow, an open source ML framework from Google, resulting in a trained model that would be ready to go into production on a desktop or server platform.

But in order to deploy this model on a microcontroller, further optimisation steps are needed to reduce required memory and computing power. These include, for example, so-called quantisation, in which inputs and parameters are represented with less precision, and pruning, in which neurons that have little impact on the final result are removed from the network. For these optimisations, we used TensorFlow Lite, another open source tool from Google, which we then converted into a C/C++ header file and were able to flash on the Raspberry Pi Pico. To interpret the model on the microcontroller, we used Google's TensorFlow Lite for Microcontrollers (TFLu) C/C++ library.

©Sirris

Results

Without any problem we managed to deploy a computer vision application on a cheap Raspberry Pi Pico microcontroller with open source tools. The model processes RGB images with a resolution of 48 x 48 pixels and classifies them into three classes with an inference time of 250 ms.

In future work, we will explore the possibilities offered by more powerful microcontrollers, with more memory and computing power than the Raspberry Pi Pico, in terms of computer vision and anomaly detection, among other things.

Webinar

On 31 January 2024, we will delve deeper into the image classification task described above on a Raspberry Pi Pico microcontroller in a free, one-hour webinar. During this webinar you may expect:

- A discussion of the hardware, microcontroller and camera;
- A discussion of the applied MobileNet architecture and comparison with alternative networks;
- A demonstration of the open source tools used for training, optimisation and inference: TensorFlow, TensorFlow Lite, TensorFlow Lite for Microcontrollers;
- A discussion of the results.

Keen to join us? Register now for our webinar!

anced by VLAIO.

Authors

Frank Van den Broek