Exploring Static Code Generation and SIMD-Acceleration for Machine Learning on RISC-V by Rafael Stahl, Technical University of Munich @RISC-V Forum on "Developer Tools & Tool Chains" on June 2, 2021 18:05 CEST

2021/05/14

The deployment of machine learning applications on microcontrollers known as TinyML enables new low-power applications and always-on devices. The RISC-V architecture is attractive for such microcontrollers, because it provides easy extensibility, a healthy ecosystem and no license costs. The major challenges with resource-constrained devices are run time and memory usage. Existing machine learning frameworks provide runtime libraries that dynamically load and execute a model, but this entails overheads. In this talk, two static code generators based on TensorFlow Lite for Microcontrollers and TVM are presented, that avoid these overheads by generating static code to execute the model. Additionally, machine learning kernel implementations based on a RISC-V version of CMSIS-NN are provided that make use of the RISC-V P- and V-Extensions to accelerate inner loops with SIMD-Operations. The contributions were evaluated on the TinyMLPerf benchmark with the ETISS simulator and show the benefits of static code generation and specialized kernel implementations (https://sched.co/jGkT [1]).

Agenda of the whole RISC-V Forum on “Developer Tools & Tool Chains” starting at 16:00 CEST: https://events.linuxfoundation.org/riscv-forum-developer-tools-and-tool-chains/program/schedule/ [2]

You have to be registered without costs for the Forum at https://community.riscv.org/events/details/risc-v-foundation-risc-v-forums-presents-risc-v-forum-developer-tools-tool-chains-1/?utm_medium=email&_hsmi=126985377&_hsenc=p2ANqtz--TfaomdDR6nbumi4lY7mhGTw1EhgwW8OEVIG_do0m7Xj6KS0Qn5j5b215aZXo4R8eer4q16BdWr0TxLd67F-IUtculPg&utm_content=126985377&utm_source=hs_email [3]

Wednesday, June 2

16:00 CEST

RISC-V Welcome - Kim McMahon, RISC-V International

16:05 CEST

RISC-V Tools & Runtime HSC Overview - Christoph Müllner, SBA Research & Philipp Tomsich, VRULL GmbH

16:15 CEST

Java on RISC-V: OpenJDK Porting Work Update - Sanhong Li & Kevin Kuai, Alibaba Cloud

16:35 CEST

Analysis for Code Size Opportunities in RISC V - Ibrahim Abu Kharmeh, Huawei UK

17:05 CEST

Programmer Productivity and Performance on Embedded RISC-V CPUs - Nick Brown, EPCC at the University of Edinburgh

17:15 CEST

CFU Playground: Model-specific Acceleration on FPGAs - Timothy Callahan & Alan V. Green,Google

17:45 CEST

Linker Relaxation in LLD - Chih-Mao Chen, Andes Technology

18:05 CEST

Exploring Static Code Generation and SIMD-Acceleration for Machine Learning on RISC-V - Rafael Stahl, Technical University of Munich

18:25 CEST

Porting and Optimization V8 for RISC-V - Ji Qiu, Institute of Software, Chinese Academy of Sciences

18:35 CEST

Panel: Toolchains & Runtime - Panelists to be Announced

Das Projekt Scale4Edge wird unter den Förderkennzeichen 16ME0122K-140, 16ME0465, 16ME0900, 16ME0901 im Förderprogramm ZuSE durch das deutsche Bundesministerium für Forschung, Technologie und Raumfahrt (BMFTR) gefördert.

Links:
[1] https://sched.co/jGkT
[2] https://events.linuxfoundation.org/riscv-forum-developer-tools-and-tool-chains/program/schedule/
[3] https://community.riscv.org/events/details/risc-v-foundation-risc-v-forums-presents-risc-v-forum-developer-tools-tool-chains-1/?utm_medium=email&_hsmi=126985377&_hsenc=p2ANqtz--TfaomdDR6nbumi4lY7mhGTw1EhgwW8OEVIG_do0m7Xj6KS0Qn5j5b215aZXo4R8eer4q16BdWr0TxLd67F-IUtculPg&utm_content=126985377&utm_source=hs_email