Verification of Model Outputs in TinyML Deployment Flow

Machine Learning on the edge is becoming more and more popular nowadays. Especially for safety critical applications it is not acceptable to have any deviation of model outputs caused by the deployment method while other application might accept differences to the golden reference outputs to some degree.

Our TVM deployment flow is mainly based on the MLonMCU (https://github.com/tum-ei-eda/mlonmcu) tool. While it already supports validating model outputs, this functionality is rather limited.

**Task Description:**

- **Goal:** Reimplement the validation feature in MLonMCU to be more flexible and accurate
- **Configuration:** In addition to bit-exact equivalences, the user may also allow deviations in some degree (absolute/relative delta)
- **Flexibility:** Provide an generic interface supporting several target-specific implementations (Offline: data compiled in ROM, Online: data sent via Semihosting, UART,...)
- **Accuracy:** Allow comparing intermediate values (between layers) in addition to just model outputs.

**Prerequisites**

- Basic Knowledge of Machine Learning
- Experience with Embedded C programming
- Good Python Coding skills
- Ideally experience using TVM Machine Learning Framework

**Contact**

Philipp van Kempen
philipp.van-kempen@tum.de

**Advisors**

Philipp van Kempen