

BA/MA-Abschlussarbeit

TinyAI Experiment: Implementierung von neuronalen Netzen auf einem FPGA Linux CPU

Kurzbeschreibung

Aufgrund ihrer inhärenten Parallelität haben sich neuronale Netze als ein sehr starkes KI-System erwiesen, das nicht nur die neuesten Entwicklungen bei großen Sprachmodellen antreibt, sondern auch ein breites Spektrum an industriellen Anwendungen. In den letzten Jahren hat sich der Schwerpunkt jedoch von großen neuronalen Modellen und stromhungrigen Anwendungen in Rechenzentren auf eingebettete und sogar sehr kleine neuronale Netzsysteme verschoben. Dieser Trend wird durch die Verbreitung eingebetteter Unterhaltungselektroniksysteme wie Fitness-Tracker, tragbare Spielgeräte usw. unterstützt. Um den Ressourcen- und Energiebedarf zu minimieren, konzentrieren sich die jüngsten Entwicklungen in der TinyAI konzentrieren sich darauf, große neuronale Netzwerke zu komprimieren und zu quantisieren, damit sie in kleine, ressourcenbeschränkte Geräte passen. Die theoretische Arbeit, die sowohl von der Industrie als auch von der akademischen Welt geleistet wird, befasst sich mit Methoden, um große Modelle mit einem angemessenen Verlust an Leistung und Genauigkeit zu verkleinern. Hier spielen einfache Implementierungen und Effizienz eine große Rolle.

In diesem Projekt geht es darum, den Einsatz neuronaler Netze auf einem kleinen Microblaze PowerPC CPU synthetisiert in eine Spartan 6 LX-9 FPGA MicroBoard Linux-Minicomputer zu untersuchen. Das kostengünstige Spartan-6 LX9 FPGA-basierte Entwicklungskit verfügt über 1.430 Slices oder 5.720 LUTs und 32 Block RAM. Umfangreiche Funktionen wie 64MB DDR, 128Mb SPI Flash, 10/100 Ethernet PHY unterstützen die Implementierung einer Microblaze PowerPC CPU, auf der ein leichtes Linux-Betriebssystem läuft. Das Ziel des Projekts ist es, ein einfaches neuronales Netzwerk für Regression und Klassifizierung zu entwickeln und einzusetzen. Das beinhaltet sowohl das Training als auch die Inferenz des leichtgewichtigen neuronalen Netzes auf dem Gerät. Ein Latenz-Benchmarking in Bezug auf die Hardwarespezifikationen soll durchgeführt werden.

Tasks

- Einführung in die Anatomie der neuronalen Netze für Regression und Klassifikation.
- Vertrautmachen mit der Spartan 6 LX-9 FPGA Linux MicroBoard Umgebung.
- Implementierung der einfachen neuronalen Netze für Regression und Klassifikation.
- Einsatz auf dem Spartan 6 LX-9 FPGA Linux MicroBoard Minicomputer und Leistungsanalyse.

Voraussetzungen

- Erfahrung mit Neuronale Netzwerke
- Gute Embedded Systems Kenntnisse
- Gute Linux Kenntnisse und sehr gute Programmierkenntnisse (C/C++, Python)

Betreuer

Prof. Dr. Ing. Cristian Axenie, M.Sc.