

I. IDENTIFIKAČNÍ ÚDAJE

Název práce:	Efficient Implementation of Neural Networks for Real-Time Applications
Jméno autora:	Matěj Suchánek
Typ práce:	bakalářská
Fakulta/ústav:	Fakulta elektrotechnická (FEL)
Katedra/ústav:	Katedra kybernetiky
Oponent práce:	Dr. rer. nat. Jan Hering
Pracoviště oponenta práce:	Katedra kybernetiky, Biomedical Imaging Algorithms (BIA)

II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

Zadání	náročnější
<i>Hodnocení náročnosti zadání závěrečné práce.</i>	
Zadání práce vyžadovalo nejen rešerši state-of-the-art metod redukce paměťové a výpočetní náročnosti při aplikaci neuronových sítí, ale i jejich implementaci a aplikaci pro experimentální ověření zrychlení a výsledné přesnosti daných modelů.	

Splnění zadání	splněno
<i>Posuďte, zda předložená závěrečná práce splňuje zadání. V komentáři případně uveďte body zadání, které nebyly zcela splněny, nebo zda je práce oproti zadání rozšířena. Nebylo-li zadání zcela splněno, pokuste se posoudit závažnost, dopady a případně i příčiny jednotlivých nedostatků.</i>	
Jednotlivé body zadání, tedy detailní rešerše aktuálních metod urychlování neuronových sítí, implementace a následně experimentální ověření vlivu přesnosti a zrychlení byly v odpovídajícím rozsahu zpracovány a výsledky dobře diskutovány.	

Zvolený postup řešení	správný
<i>Posuďte, zda student zvolil správný postup nebo metody řešení.</i>	
Zadání práce klade důraz na experimentální ověření, výsledky experimentů v naprosté většině případů korespondují s teoretickými předpoklady. Student musel při experimentech vyřešit i problémy s kompatibilitou mezi GPU hardwarem a dostupnými verzemi jednotlivých frameworků. Jako drobné negativum hodnotím, že u experimentu s efektivními architekturami (3.4) bylo jako srovnávací GPU zvolena desktopová Titan X a ne Maxwell GPU na platformě Jetson, když se v tomto experimentu vyhodnocuje ARM CPU na Jetson Nano.	

Odborná úroveň	A - výborně
<i>Posuďte úroveň odbornosti závěrečné práce, využití znalostí získaných studiem a z odborné literatury, využití podkladů a dat získaných z praxe.</i>	
Téma real-time aplikací neuronových sítí je téma velmi aktuální. Prezentované experimenty dobře dokládají možnosti zrychlování aplikace neuronových sítí pro tři zásadní kategorie – klasifikaci (kategorizaci), detekci objektů a sémantickou segmentaci. Součástí práce jsou nejen implementované skripty pro kvantizaci, ale také aplikace sítí na moderní, AI-specifické výpočetní platformě Jetson Nano.	

Formální a jazyková úroveň, rozsah práce	A - výborně
<i>Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.</i>	
Text práce je velmi čtivý, a to díky vysoké jazykové úrovni a díky vhodné strukturovanosti textu – kapitoly jsou uvozené souhrnným přehledem, čtenář má tak odpovídající představu o výstavbě a návaznosti jednotlivých kapitol a sekcí. Typografická úprava i rozsah práce jsou odpovídající, formální zápisy jsou v pořádku.	

Výběr zdrojů, korektnost citací

B - velmi dobře

Vyjádřete se k aktivitě studenta při získávání a využívání studijních materiálů k řešení závěrečné práce. Charakterizujte výběr pramenů. Posuďte, zda student využil všechny relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků a úvah, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami.

Formát bibliografických citací je v pořádku, použité prameny jsou aktuální, což je u daného tématu nutností. Po citační stránce shledávám jako lehce nestandardní, že je sekce *Symmetric quantization* postavena převážně na jedné prezentaci (citace [18]).

Další komentáře a hodnocení

Vyjádřete se k úrovni dosažených hlavních výsledků závěrečné práce, např. k úrovni teoretických výsledků, nebo k úrovni funkčnosti technického nebo programového vytvořeného řešení, publikačním výstupům, experimentální zručnosti apod.

III. CELKOVÉ HODNOCENÍ, OTÁZKY K OBHAJOBĚ, NÁVRH KLASIFIKACE

Shrňte aspekty závěrečné práce, které nejvíce ovlivnily Vaše celkové hodnocení. Uveďte případné otázky, které by měl student zodpovědět při obhajobě závěrečné práce před komisí.

Předloženou bakalářskou práci hodnotím jako výbornou. Student obsáhl formou řešerše aktuální metody a softwarové nástrojů pro efektivní implementaci (aplikaci) neuronových sítí. Tyto teoretické zkušenosti pak převedl do praxe v rozsáhlé sadě experimentů, a to na odlišných hardwarových platformách. Pomocí provedených experimentů student velmi dobře ilustruje a také diskutuje přínosy ve zrychlení a případnou cenu tohoto zrychlení na úkor přesnosti.

Doplňující otázky

1. Bylo by možné dosáhnout dalšího zrychlení aplikováním kvantizace na moderní (tenké) architektury typu ShuffleNet nebo MobileNet?

Předloženou závěrečnou práci hodnotím klasifikačním stupněm **A - výborně**.

Datum: 24.8.2020

Podpis: