

I. IDENTIFIKAČNÍ ÚDAJE

Název práce:	Adding Context into Convolutional Neural Networks
Jméno autora:	Sebastian Štefko
Typ práce:	bakalářská
Fakulta/ústav:	Fakulta elektrotechnická (FEL)
Katedra/ústav:	Katedra kybernetiky
Oponent práce:	Karel Zimmermann
Pracoviště oponenta práce:	Katedra kybernetiky

II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

Zadání	průměrně náročné
<i>Hodnocení náročnosti zadání závěrečné práce.</i>	
Vložte komentář.	

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>	
Vložte komentář.	

Zvolený postup řešení	správný
<i>Posuďte, zda student zvolil správný postup nebo metody řešení.</i>	
Vložte komentář.	

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>	
Student naimplementoval několik různých technik pro vkládání kontextu do hlubokých konvolučních sítí. Výsledky velmi dobře vyhodnotil a výsledky detailně diskutoval.	

Formální a jazyková úroveň, rozsah práce	B - velmi dobře
<i>Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.</i>	
Práce je dobře čitelná, a velmi pozitivně hodnotím diskuzi za každou podkapitolou experimentů. Formální zápisy jsou místy nepřesné (např. u různých názvů „mean attention distance“, „ERF mean distance“, „mean distance“ není explicitně vysvětleno, jak spolu souvisí). Také mi chybí popis NextConv, která je použita v experimentální části.	
Drobnosti:	
Kapitola 1:	
Obr.1.2. z textu práce není jasné, co přesně znamená „mean attention distance“ (je mean vážený attention? Hraje nějakou roli gradient? je mean přes více obrázků?)	
Kapitola 4.1:	
<ul style="list-style-type: none"> • Je to ta samá „mean distance“, která je použita v Obr 1.2 – asi ne, ale pokud ano asi by chtělo přidat odkaz na tuto kapitolu. • F je matice, $x \in F$ se myslí jeden element z matice (tj. skalár), a pak zavedete Eukleidovskou vzdálenost od skaláru, ale ve skutečnosti myslíte vzdálenost od souřadnic tohoto elementu. Technicky je jasné, ale matematicky byste měl zapsat trochu lépe. 	

- Je průměrováno přes 10 obrázků dostatečně vypovídající? Jaká ji minimální a maximální hodnota na validačních datech?

Kapitola 3.1:

- Str 13: Předpokládám, že se uvažuje pouze prostorové sousedství a že náhodné kontextové permutace jsou náhodně vybrány při vytváření architektury, ale v průběhu trénování (a inference) zůstávají zafixované. Z jak velkého sousedství jsou náhodné permutace generovány? Jsou samplovány rovnoměrně?
- Obr 3.1 vytváří dojem že jednotlivé příznaky nejsou organizovány v příznakové mapě, ale mohou být volně umístěny v prostoru, což asi není pravda.

Výběr zdrojů, korektnost citací

A - výborně

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.

Vložte komentář.

Další komentáře a hodnocení

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

Vložte komentář (nepovinné hodnocení).

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.

Výsledkem práce je několik architektur, které rozšiřují RF/ERF pro ResNet50. Navržené architektury jsou porovnány se state-of-the-art sítěmi jakou jsou HRNet, NextConv nebo ViT. Experimenty pěkně demonstrují vliv architektury na top-1 accuracy, RF/ERF a mean attention distance. Celkově bylo odvedeno hodně experimentální práce, ale vlastní text by šlo ještě zlepšit (exaktní definice RF, ERF, mean attention distance není uvedena).

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

Otázky do diskuze:

1. Proč selhal VisionTransformer na ImageNet datasetu? ViT má nejvyšší počet parametrů, je selhání způsobeno přefitováním nebo neschopností minimalizovat chybu na trénovacích příkladech? Jaké byly trénovací chyby?
2. Obr.1.3. ukazuje, že RF prostředního příznaku v druhé vrstvě je 100×100 = je to správně? Z architektury uvedené na Obr.2.1 to vypadá, že první i druhá vrstva mají stride 2, dale stride není uveden. Jak přesně se RF spočetl?
3. Tabulka 4.5 porovnává výsledky jednotlivých architektur s dilatovanou konvolucí. Až na modified dilated je rozdíl v Top-1 accuracy relativně malý (okolo 0.5%) vzhledem k původní architektuře. Jsou rozdíly mezi jednotlivými architekturami konzistentní i při opakování experimentů (např s jinou inicializací nebo hyperparametry)?
4. Kapitola 3.3. str 15-16. Self-attention zvyšuje váhy permutacím, které jsou podobné originální části příznakové mapy. Jaké jsou reálné příklady, kdy tento přístup bude fungovat dobře a kdy špatně? Nešlo by generování náhodných permutací vychýlit, tak aby pravděpodobnost podobných permutací byla vyšší? Dávalo by to smysl?

Datum: 1.5.2022

Podpis: