

Review of Master's Thesis

Student: Kvita Jakub, Bc.
Title: Image Captioning with Recurrent Neural Networks (id 18734)
Reviewer: Španěl Michal, Ing., Ph.D., UPGM FIT VUT

- 1. Assignment complexity** **considerably demanding assignment**
Zadání se věnuje metodám generování popisů fotografií s využitím konvolučních a rekurentních sítí, což je velmi aktuální problematika se značným výzkumným potenciálem, kterou se výzkumníci zabývají až v posledních letech. Kromě současných přístupů, musel autor nastudovat i množství pokročilých metod, jako samotný základ pro svoji práci (CNN, RNN, jazykové modely, atd.).
- 2. Completeness of assignment requirements** **assignment fulfilled**
- 3. Length of technical report** **in usual extent**
Autor zvolil rozumnou úroveň rozboru problematiky a převzaté teorie týkající se tématu.
- 4. Presentation level of technical report** **75 p. (C)**
Technická zpráva je přehledná a je dobře pochopitelné co a jak autor tvořil. Popis současných metod pro generování popisů fotografií je zpracovaný velmi dobře a pokrývá zásadní publikace posledních let, ale ne vždy je úplně srozumitelný. Relativně všeobecný popis se střídá s detaily, které nejsou úplně dovysvětleny.
 - Za nevhodné považuji tři strany Implementace věnované přehledu použitelných existujících knihoven a nástrojů, z nichž autor nakonec použil pouze Torch.
- 5. Formal aspects of technical report** **60 p. (D)**
Zpráva je psaná v angličtině a její úroveň je průměrná. Text je sice srozumitelný, ale dojem kazí poměrně dost elementárních chyb v čárkách, členech, občas i ve struktuře složitějších vět. Příkladem budiž samotný abstrakt. V názvech kapitol chybí velká písmena. Autor jednou použije korektně "in Figure", ale jindy "on Figure" nebo "on the Figure"...
 - Odkazy do literatury se u obrázků objevují až za koncem věty ("model. [54]").
 - Ne všechny symboly použité v rovnicích jsou popsány.
- 6. Literature usage** **80 p. (B)**
Seznam literatury čítá úctyhodných 59 publikací. Z řešení je patrné, že autor mnoho z nich opravdu prostudoval a nasál potřebné znalosti, ale je až těžko uvěřitelné, že prošel opravdu všechny citované zdroje. Převzaté prvky jsou řádně odděleny od vlastních, pouze některé obrázky se zdají být převzaté, ale zdroj uveden není.
- 7. Implementation results** **75 p. (C)**
Výstupem práce jsou zejména výsledky experimentů s navrženou architekturou sítě a podpůrné skripty pro práci s datovou sadou, trénování sítě, atd. Uvedené je realizováno na dobré úrovni. Na přiloženém CD chybí popis obsahu a najdeme zde pouze zdrojové kódy skriptů v Lua, video a to je vše. Žádný příklad, ukázka datasetu, apod.
- 8. Utilizability of results**
Nejedná se o práci pouhého kompilačního charakteru. Autor navrhnul a otestoval vlastní řešení pro popis fotografií, které bohužel nedává uspokojivé výsledky. Jako zásadní problém vidím použití předtrénované konvoluční sítě pro klasifikaci obrázků, která se typicky snaží v obraze hledat smysluplné struktury popsatelné "na úrovni slov", ale jazykový model, v kontrastu s publikovanými pracemi, stavěl na úrovni jednotlivých znaků...
- 9. Questions for defence**
 - V citovaných článcích [54, 57] autoři používají RNN pro predikci slov a nikoliv znaků. Na základě čeho jste usoudil, že predikce na úrovni znaků může fungovat? Nesouvisí i vznik nesmyslných slov s predikcí popisků na úrovni znaků?
 - Výstup CNN je také typicky vstupem RNN části, co Vás vedlo k rozhodnutí výstup CNN přivést jako vnitřní stav?
 - Vysvětlíte, jak v případě BoW experimentu inicializujete LSTM, které jsou určeny pro predikci na úrovni znaků?
- 10. Total assessment** **80 p. very good (B)**
Pan Kvita se zhostil náročného zadání se ctí a je patrné, že na řešení usilovně pracoval. Bohužel se zdá, že se z

nastudované literatury úplně nepoučil a vydal se vlastní cestou, která však nebyla úspěšná. Přestože prezentované výsledky nejsou dobré a technická zpráva mohla být zpracována lépe, hodnotím celkové zpracování obtížného tématu diplomové práce jako nadprůměrné a překračující běžné práce.

In Brno 10. June 2016

.....
signature