Posudek oponenta diplomové práce

Autor/autorka práce: **Václav Honzík**

Název práce: **Multi-modální zpracování dokumentů**

Práce se zabývá multi-modálním zpracováním obrazových dokumentů, konkrétně skenů německy psaných vlastivěd z portálu Porta fontium. Cílem je analýza rozložení stránek, tj. zejména detekce textových bloků a jejich klasifikace do tříd (odstavec, nadpis apod.). Pro analýzu rozložení jsou využity jak obrazový vstup, tak i textový obsah (získaný pomocí OCR nástroje).

V teoretické části jsou nejprve shrnuty metody pro zpracování obrazu a textu pomocí neuronových sítí. Následuje přehled metod využívajících multi-modální vstup a také přehled dostupných datových sad. Dále je uveden přehled OCR nástrojů a provedeno zhodnocení jejich použitelnosti pro danou úlohu.

V rámci práce byla vytvořena nová datová sada „Heimatkunde“, která byla autorem anotována pro úlohu rozpoznání rozložení stránky. Anotace zahrnují 7 tříd textových bloků. Z popisu není zcela patrné, jestli byly anotovány původní skeny (dvoustrany), nebo jestli byly obrázky nejprve rozděleny na jednotlivé stránky a anotovány zvlášť. Část datasetu je také anotována pro možnost trénování OCR. Kapitola popisující dataset obsahuje rovněž experimenty s trénováním OCR modelů, které by bylo vhodnější umístit až k ostatním experimentům.

V kapitole 7 je návrh řešení multi-modálního klasifikátoru, který kombinuje segmentační a klasifikační model. Výsledky obou modelů i výsledného multi-modálního modelu jsou v kapitole 8. V kapitole 8 je uvedena celá řada výsledků, ve kterých je poměrně obtížné se orientovat. Bylo by vhodné přidat shrnující tabulku, která porovná multi-modální metody s obrázkovými a umožní čtenáři lepší srovnání, případně kapitolu samostatně zhodnotit.

Přiložený software je funkční a dostatečně popsaný i komentovaný. Instalace a spuštění podle přiložených „readme“ souborů je bezproblémové.

Text práce je psaný anglicky, což hodnotím jako dobrou volbu, zejména z důvodu problematického překladu některých termínů do češtiny. Angličtina je na slušné úrovni, občasné drobné chyby (členy, záměna příslovce a přídavného jména) nepředstavují problém. Celkově je struktura práce vhodná a přehledná, až na výjimky uvedené výše. Zejména kapitola 8 by mohla být přehlednější.

Práci hodnotím jako rozsahově nadstandardní. Bylo provedeno velké množství experimentů, které vyžadovaly velký objem výpočetního času a výsledky přinášejí zajímavé srovnání různých architektur neuronových sítí a jejich kombinací.

Zadání bylo splněno bez výhrad.

Dotazy k práci

1) U popisu datasetu píšete, že anotace pro OCR probíhala na řádcích, popřípadě blocích textu (str. 41). Byly opravdu použity i anotace bloků textu?

2) U popisu datasetu zmiňujete, že byly vytvořeny trénovací a testovací sady. Jak jste řešil validační data při trénování modelů?

3) Tabulka 8.9 ukazuje experiment s použitím uměle poškozených textů. Z výsledků vyplývá, že textový vstup má pouze malý vliv na celkový výsledek multi-modální klasifikace. Zkoušel jste klasifikovat bloky textu pouze na základě textové reprezentace?

4) Vysvětlete proměnnou D_k v rovnici 2.2.

Navrhuji hodnocení známkou **výborně** a práci doporučuji k obhajobě.

V Plzni 5.6.2023 Ing. Ladislav Lenc, Ph.D.