

Posudek diplomové práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce Bc. Štěpán Procházka

Název práce Adaptive Handwritten Text Recognition

Rok odevzdání 2021

Studijní program Informatika **Studijní obor** Umělá inteligence

Autor posudku Pavel Straňák **Role** oponent

Pracoviště Ústav formální a aplikované lingvistiky, MFF UK

Text posudku:

Práce *Adaptive Handwritten Text Recognition* je napsána vynikající angličtinou s pouze výjimečnými chybami, které až na jednu nebo dvě výjimky neztěžují chápání textu.

V abstraktu a následně v úvodní kapitole Introduction student představí motivaci pro digitalizaci dokumentů, jejíž součástí je OCR, jehož podmnožinou je pak OCR rukopisů. V tomto představení se bohužel nevyhne několika nepřesnostem a nedůslednostem: 1. V abstraktu motivačně uvádí, že rozpoznávání psaného textu je úloha, které se nevěnuje pozornost. Jednak to není pravda v oblasti, kterou práce zcela pomíjí, totiž interaktivní vstup textu na mobilních zařízeních, zadruhé to není pravda v posledních letech, ale autor tyto práce pomíjí (viz níže). 2. OCR opravdu není celý proces digitalizace, tedy jak autor píše “conversion of physical documents to their digital counterparts.” Konverzí fyzického artefaktu do digitální podoby se naopak OCR – ani předložená práce – nezabývá, jen využívá její výsledky, tedy digitální obrazová data. 3. Popis OCR systémů jako celku od analýzy obrazu přes detekci a rozpoznání textu až po uchování maxima informací vč. vzhledu a smyslu částí obrazu a textů ve strukturovaných formátech v práci není. Úvod a motivaci až po sekci 1.1 Text Detection, která se omezí na to, že problém existuje, tedy považují za poněkud odbyté, což je škoda.

Od sekce 1.2, kdy se práce již koncentruje na samotné rozpoznání textu v rámci řádku, se podobné nepřesnosti již nevyskytují. Co ale postrádám, je větší množství citací. Když se autor rozhodl v sekci 1.2.2 Connectionist Temporal Classification popsat použité algoritmy vč. například beam search pro decoding, tak u nich nikde neuvádí citace. Nezabývá mi než předpokládat, že vše tedy uvádí podle citace Graves et al. (2012) v úvodu sekce, ale jasné to není.

V sekci 1.3 Language Modelling platí stejná výtka ohledně citací, ale ještě výrazněji. U věty začínající “Historically, rule based models based on expert knowledge were used” mě opravdu zajímá odkaz na ručně tvořené jazykové modely, které neznám. Bohužel tam není. Následně autor zmíní vynález n-gramového jazykového modelu, opět zcela bez citace, i když jim pak věnuje celou

podsekci.

Následuje kapitola o syntetických datech, v níž je i sekce o kurzivním písmu (fonty napodobující nebo předepisující rukopis) a jeho sazbě. Zde je zajímavá zmínka o fontech užitých pro syntézu, vč. jednoho specificky českého, který implementuje českou psanou školní abecedu známou z 80. let. Ostatní uvedené fonty jsou podle ukázek založeny na jiných psaných stylech, převážně asi anglosaských, což je zřetelné hlavně ve verzálcích nebo také ve zcela odlišném malém “z”. Bylo by zajímavé se věnovat více tomuto aspektu, hlavně zda a jak pomáhá český font u dat z amerického prostředí a naopak, vzhledem ke v práci použitým českým a americkým datům.

V sekci 3 Our Solution, kde dojde na specifika vlastního řešení, je příliš stručný popis vlastního visuals feature extractor. “volně se inspirujeme systémem Shi et al. 2015, ale architekturu jsme vylepšili těmito 4 změnami: strided convolutions namísto max-pooling layers, atd.” je popis příliš stručný, Motivace i evaluace změn zcela chybí. Považuji to za politováníhodné, obzvláště když jde o jádro vlastní práce a srovnáme to s popisy standardních algoritmů na jiných místech textu. Drobným nedostatkem je v sekci 3 a 4 také to, že tabulky ani jejich popisky neobsahují označení, v jakých jednotkách a co znamenají čísla v nich. Jistě to je možno dohledat z kontextu v textu u odkazu na tabulku, ale opět to není dobrá praxe.

K sekci 3.2.1 mám otázku na užití fontu Abeceda: tento font vyžaduje pro plynulé navazování znaků vkládání speciálních znaků, glyfů neznačících písmena abecedy, mezi znaky. Bylo by zajímavé vědět, jak to student řešil, nebo zda to není pro výslednou kvalitu syntetických dat potřebné.

V tabulce 4.1 není uvedena reference při srovnání s publikovanými výsledky a reference v textu není jednoznačná. Přesnost ve srovnání s předchozími výsledky je zásadně důležitá.

Tím se dostávám k největší výtce k předložené práci: nedostatečné práci s literaturou a srovnání s dosaženými výsledky. Jak jsem již zmínil výše, hned v úvodu práce se uvádí, že rozpoznávání rukopisu je problém, kterému je dosud věnována malá pozornost. Toto tvrzení považuji za poněkud překvapivé. Jednak zde máme systémy pro vstup textu rukopisem místo klávesnice. Od prvního nasazení v kapesním zařízení Newton (<http://shinyverse.org/larry/ANHR.html>) po integraci rozpoznávání rukopisu v iPadOS 14 je souvislý rukopis podporován v mobilních zařízeních. Z pochopitelných důvodů je ještě důležitější než pro latinkové jazyky pro jazyky dálného východu, kde je na mobilních zařízeních vstup textu psáním zcela rutinní a obsahují je všechny naše dnešní mobilní telefony. Z druhé strany i kdybychom rozpoznání textu omezili jen na psaní na analogové medium a následnou digitalizaci, zdá se, že problém není zrovna neřešený: Google (<https://cloud.google.com/vision/docs/ocr>), Microsoft (<https://docs.microsoft.com/en-us/azure/cognitive-services/computer-vision/quickstarts-sdk/client-library>), i řada dalších firem (<https://www.hanvon.com/show-8-1-1.html>) poskytuje svá řešení (<https://research.aimultiple.com/handwriting-recognition/>, <https://www.hanvon.com/show-8-1-1>).

html) Stejně tak akademické projekty řeší explicitně OCR historického rukopisu: (<https://readcoop.eu/transkribus/?sc=Transkribus>). Dokonce i pro češtinu existuje projekt, který řeší specificky rozpoznání českých rukopisných textů: Projekt Pero řešený VUT a MZK (<https://pero.fit.vutbr.cz>). Práce autorů tohoto řešení je v předložené DP i citována (Kiss et al. 2021), ale pouze specificky ohledně self-trainingu a ani zde není žádné bližší srovnání. Bez popisu state of the art v oboru a bez srovnání s ním je pro mě jako recenzenta nemožné říci, jak jsou výsledky v předložené práci kvalitní a důležité.

Práce na mě působí dojmem, že její nejdůležitější část je v sekci 4.3.2 Self-Training. Předložené výsledky vypadají opravdu zajímavě. Obzvláště střídání dekodérů v jednotlivých iteracích a výsledky takto dosažené působí velmi přínosně. Zda opravdu ale tak zajímavé jsou, to nemohu říci z důvodu chybějícího srovnání.

Práce na mě tedy ve výsledku působí rozporuplně. Je zjevné, že student odvedl velké množství práce, mimo jiné vytvořil a zveřejnil kvalitní a důležité datasety. Pokud jde o oblast OCR rukopisu celkově, mám řadu výhrad, ale pokud práce měla být primárně o unsupervised trénování, daly by se ony výhrady z velké části shrnout do konstatování, že měla práce být napsána tak, aby její cíl a zaměření byly zřejmé. I pak by bylo velmi vhodné srovnání přinejmenším se systémem Pero. I zde je možné, že nejde o zásadní opomenutí, jelikož jde o systém velmi nový a některé výsledky publikují až letos, tedy paralelně se vznikem předložené práce: <https://pero.fit.vutbr.cz/publications>.

Práci rozhodně doporučuji k obhajobě, z jejíhož průběhu by mělo vyplynout, jaký je její opravdový přínos.

Práci doporučuji k obhajobě.

Práci nenavrhuji na zvláštní ocenění.

V Praze dne 17. srpna 2021

Podpis: