

# Posudek bakalářské práce

Matematicko-fyzikální fakulta Univerzity Karlovy

**Autor práce** Kristýna Harvanová

**Název práce** Postprocessing syntetických notopisů v kontextu jejich rozpoznávání

**Rok odevzdání** 2024

**Studijní program** Informatika

**Specializace** Umělá inteligence

**Autor posudku** Jan Hajič jr.

**Role** Oponent

**Pracoviště** Ústav formální a aplikované lingvistiky

Prosím vyplňte hodnocení křížkem u každého kritéria. Hodnocení *OK* označuje práci, která kritérium vhodným způsobem splňuje. Hodnocení *lepší* a *horší* označují splnění nad a pod rámec obvyklý pro bakalářskou práci, hodnocení *nevyhovuje* označuje práci, která by neměla být obhájena. Hodnocení v případě potřeby doplňte komentářem. Komentář prosím doplňte všude, kde je hodnocení jiné než *OK*.

K celé práci	lepší	OK	horší	nevyhovuje
Obtížnost zadání	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Splnění zadání	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Rozsah práce ... <i>textová i implementační část, zohlednění náročnosti</i>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Práce se zabývá převedením černobílého syntetizovaného obrazu notopisu na více realistický obrázek podobnější skutečným notovým digitalizátům. Jedná se pro rozpoznávání notopisu (OMR) o důležitou úlohu, neboť trénovacích dat prakticky pro libovolný přístup k OMR je nedostatek. Autorka implementuje především syntézu pozadí pomocí metody image quilting, kde se pozadí postupně sestavuje z částí jiných obrázků příslušné textury. Jednotlivá vybraná pozadí jsou zvolená velmi reprezentativně. Výsledek (obr. 3.3) vypadá realisticky a na základě výsledků pokusů v části 5.3 syntéza (nepřekvapivě, ale přesvědčivě) vylepšuje detekci taktů. Další implementované metody jsou simulace průsaku inkoustu z druhé strany papíru (častý fenomén u notových archiválií), simulace rukopisu, resp. pera, pomocí morfologické dilatace a eroze, a kanungo šum. Nedělá se však barevný model inkoustu a rozpíjení.

Pokusy jsou navrženy dobře, byť se ne všechny podařily (a zde oceňuji, že neúspěšné trénování detekce notových hlaviček bylo v práci řádně diskutováno, spíše než zamlčeno). Oceňuji použití jak tištěných, tak rukopisných not pro evaluaci, a zároveň trénování pouze na syntetických obrázcích. (V tomto případě není ještě k dispozici dostatek anotovaných "skutečných" dat na to, aby se daly dělat jinak obvyklé pokusy ukazující přínos syntetických dat při různých objemech trénovacích dat "skutečných".) Autorka také šikovně získala anotace bounding boxes pro jednotlivé objekty přímo z MuseScore. Ablační analýza je taktéž správný postup, nicméně mohla být ještě doplněná o opačný postup — přidávat metody degradace jednotlivě a vidět tak jejich sílu zvlášť. (Tab. 5.2 naznačuje, že pro detekci taktů by nejlépejší ve skutečnosti mohla být samotná syntéza textury pozadí: odstranění libovolné jiné metody způsobí, že je výsledek lepší, než když se metody použijí všechny. Na druhou stranu pro osnovy se zdá být nejdůležitější kanungo šum.) Každopádně výsledky ukazují, že navržený a implementovaný postup může být přínosný.

Slabší stránkou je přehled příbuzných prací, takže není zcela zřejmé, do jaké míry je tento postup přínosný oproti alternativním systémům. Práce totiž nezahrnuje diskusi těchto metod používaných pro dokumenty obecně, a s existujícími nástroji se vůbec nepracuje (DocCreator <https://doc-creator.labri.fr/> je sice nevhodný, protože nemá snadno ovladatelné API, ale existují knihovny jako <https://github.com/albumentations-team/albumentations>, <https://github.com/aleju/imgaug>, a nejbližší možná <https://github.com/sparkfish/augraphy>). Je také řada dílčích prací a materiálů, např. pro syntézu stínu (<https://www.mdpi.com/1424-8220/24/2/654>), nebo syntetická data pro trénování čtečky řídičských průkazů (<https://python.plainenglish.io/how-to-create-synthetic-datasets-of-document-images-5f140dee5e40?gi=a9f8a4ce1907>). Ne vše lze pro syntézu realistického notopisu použít, ale takové nástroje by měly být alespoň zmíněny a jejich vhodnost či nevhodnost diskutována. Nezmiňuje se také vůbec např. existence diffusion-based modelů a dalších přístupů založených na hlubokém učení (například pro notopis CycleGAN - Shatri et al., WoRMS 2023, nebo diplomová práce <https://github.com/rebelgiri/thesis-domain-adaptation-of-synthetic-document-images-using-neural-networks>). Není třeba tyto postupy používat — pro danou úlohu je to "s kanonem na vrabce" — ale měly by být zmíněny. Je možné, že by se aplikací existujícího nástroje dosáhlo s menším množstvím práce lepších výsledků (a stihlo by se více pokusů), a pokud ne, tak by pro vývoj OMR bylo žádoucí vědět proč.

Celkově však nárokům kladeným na bakalářské práce autorka jednoznačně vyhověla.

### Textová část práce

lepší OK horší nevyhovuje

Formální úprava ... jazyková úroveň, typografická úroveň, citace

Struktura textu ... kontext, cíle, analýza, návrh, vyhodnocení, úroveň detailu

Analýza

Vývojová dokumentace

Uživatelská dokumentace

Práce je psaná jasně, bez chyb, má logickou strukturu a splňuje všechny náležitosti odborného stylu. Bibliografie by mohla být rozsáhlejší (viz výše), nicméně stále je adekvátní. V podstatě není co komentovat, až na několik drobností:

Prosak —> průsak. (Jak byla volena intenzita průsaku?)

S.7: "na fakultě" - "na Matematicko-fyzikální fakultě UK"

S.21: "Oba procesy dohromady dodávají obrázku realistické vzezření, které odráží nepravidelnost a plynulost lidského rukopisu."

Vhodnější: "částečně napodobují..." (Je v kaligrafické deformaci nějaká randomizace?)

S.31: "To potvrdilo nesmírné rozdíly s notopisy tištěnými" — "...rozdíly oproti notopisům..."

S.33 Dole: "...ale jeho predikce jsou mnohonásobně přesnější." To z Obr. 5.1 není tak jednoznačně vidět. Viditelná čísla ukazují vyšší jistotu, ale není zřejmé, jestli je celková přesnost lepší — to by se mělo změřit.

S.34: Není jasné, jaký je vztah naměřených hodnot k předchozímu — spíše žádný?

S.37 nahoře: "Tento stav již není notopisem..." To sice není (validním) notopisem, ale na úlohu detekce objektů nemá validita zobrazeného notopisu vliv a v žádné ze zmiňovaných metod také nefiguruje, takže by to neměl být problém. Jinak souhlasím, že "stříhání" obrázku je vhodný nápad.

Ad dokumentace: viz poznámky k implementaci.

<b>Implementační část práce</b>	lepší	OK	horší	nevyhovuje
Kvalita návrhu ... <i>architektura, struktury a algoritmy, použité technologie</i>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Kvalita zpracování ... <i>jmenné konvence, formátování, komentáře, testování</i>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Stabilita implementace	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Zdrojový kód je (jako většina softwaru pro pokusy) užitečné minimum, které umožňuje práci replikovat. Instalační pokyny jsou správně (a jsou využity standardní postupy pro balíčkování Pythonu). README poskytuje přehled o tom, co kde v repozitáři je, a také dobře dokumentuje formáty dat. Funkci vývojové dokumentace (tedy přehledu o architektuře celého systému, byť malého) však neplní: v přehledu struktury adresáře je často pouze `validation/`: `Scripts for validation dataset.` — přičemž ani z textu práce, ani z ničeho jiného není zřejmé, co mají dělat a proč. V kódu pro evaluaci se pracuje s pojmem `workbench`, jehož role není nikde vysvětlena. Docstringy jsou minimální.

Větší péče měla být věnována také začištění skriptů. V jednotlivých modulech jsou dostupné jako `__main__.py`, což ovšem (zatím) není obvyklý postup — bylo by vhodnější je mít pojmenované, anebo mít v README postup, jak přesně spustit který pokus (což chybí). Také se jednotlivé pokusy ablační analýzy pouští komentováním a odkomentováním částí skriptu, což by mělo být ošetřeno parametry na příkazové řádce a/nebo konfiguračními soubory.

Zdrojový kód samotný je však poměrně přehledný a rozumně strukturovaný, komentáře obsahuje a naopak se vyhýbá špagetovému kódu.

Softwarová část je z definice práce spíše podpůrný materiál než její jádro, takže tyto nedostatky nejsou pro celou práci zásadní, nicméně spolu s neúplným přehledem příbuzných prací jsou nedostatky v dokumentaci důvod navrhnout `”velmi dobře”` spíše než `”výborně”`.

**Celkové hodnocení** Velmi dobře

**Práci navrhuji na zvláštní ocenění** Ne

**Datum** 25. 8. 2024

**Podpis**