

I. IDENTIFIKAČNÍ ÚDAJE

Název práce:	News Article Layout Extraction from Bitmaps Files
Jméno autora:	Vít Zeman
Typ práce:	bakalářská
Fakulta/ústav:	Fakulta elektrotechnická (FEL)
Katedra/ústav:	Katedra řídicí techniky
Vedoucí práce:	Ing. Jan Drchal, PhD.
Pracoviště vedoucího práce:	Katedra počítačů

II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

Zadání	průměrně náročně
<i>Hodnocení náročnosti zadání závěrečné práce.</i>	
<p>Extrakce rozložení stránky (layout extraction) je dobře zpracované téma. Existuje řada existujících nástrojů a knihoven, které je možnost využít (v našem případě jsme použili knihovnu detectron2). Na druhou stranu jsou existující state-of-the-art metody komplexní a jejich zvládnutí je netriviální. Těžiště práce vidím v předzpracování dat, v plánování experimentů a analýze jejich výsledků.</p>	

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>	
<p>Zadání bylo bez výhrad splněno.</p>	

Aktivita a samostatnost při zpracování práce	A - výborně
<i>Posuďte, zda byl student během řešení aktivní, zda dodržoval dohodnuté termíny, jestli své řešení průběžně konzultoval a zda byl na konzultace dostatečně připraven. Posuďte schopnost studenta samostatně tvůrčí práce.</i>	
<p>Spolupráce se studentem byla bezchybná. Termíny bez problémů dodržoval a na konzultace byl vždy dobře připraven. Student byl samostatný, jak při hledání relevantních zdrojů, tak při implementaci.</p>	

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>	
<p>K odborné úrovni nemám výhrady. Student naplánoval a provedl řadu experimentů se zajímavými výsledky. Práci mu ztěžovalo pozdější dodání vývojové verze nového datasetu VBAS konstruovaného v rámci závěrečné práce jiného studenta. To a vytíženost výpočetního clusteru vedly k tomu, že jsme nebyli schopni spustit další návazné experimenty - zejména ty, experimentující s ještě menším a nulovým zastoupením trénovacích dat (odvozené datasety PrX).</p>	

Formální a jazyková úroveň, rozsah práce	C - dobře
<i>Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.</i>	
<p>Práce je psána kvalitní angličtinou s občasnými překlepy. Rozsah je naprosto vyhovující. Práce je celkově dobře strukturována. Jediným závažnějším problémem pro čtenáře, může být nedostatečné oddělení popisu zdrojových dat a odvozených datasetů. Tato část by si zasloužila o něco podrobnější popis – např. z popisu tabulky 3.1 není jasná motivace pro odvozený dataset VBAS1. Více textu by si zasloužila i sekce 5.1 týkající se předtrénování segmentačního modelu. K formálnímu zápisu nemám výhrady.</p>	

Výběr zdrojů, korektnost citací	A - výborně
<i>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</i>	

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.

Zdroje jsou citovány správně.

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.

Výstupy této práce a naučené modely plánujeme využít v rámci většího projektu, jenž se orientuje na segmentaci článků a klasifikaci elementů stran na základě vektorové reprezentace (tiskových předloh ve formátu PDF). Předpokládáme, že výsledná metoda bude těžit z kombinace rozpoznávání nad rasterovou reprezentací.

III. CELKOVÉ HODNOCENÍ A NÁVRH KLASIFIKACE

Shrňte aspekty závěrečné práce, které nejvíce ovlivnily Vaše celkové hodnocení.

I přes menší výhrady k samotnému textu, oceňuji velký objem kvalitně odvedené implementační a experimentální práce a předloženou závěrečnou práci hodnotím klasifikačním stupněm **A - výborně**.

Datum: 1.6.2021

Podpis:

I. IDENTIFIKAČNÍ ÚDAJE

Název práce:	Extrakce rozložení novinových článků z rasterových předloh
Jméno autora:	Vít Zeman
Typ práce:	bakalářská
Fakulta/ústav:	Fakulta elektrotechnická (FEL)
Katedra/ústav:	Katedra řídicí techniky
Oponent práce:	Ing. Jiří Kubalík, Ph.D.
Pracoviště oponenta práce:	CIIRC, ČVUT

II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

Zadání	průměrně náročné
<i>Hodnocení náročnosti zadání závěrečné práce.</i>	
Tato práce se zabývá extrakcí novinových článků z rasterových předloh. Cílem bylo zpracovat přehled moderních metod pro extrakci článků a vybrané metody naimplementovat a otestovat na několika datasetech. Jedná se o standardní zadání o průměrné náročnosti.	

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>	
Student splnil zadání ve všech bodech.	

Zvolený postup řešení	správný
<i>Posuďte, zda student zvolil správný postup nebo metody řešení.</i>	
Student postupoval přesně dle zadání. Nastudoval relevantní literaturu se zaměřením na metody založené na konvolučních neuronových sítích. Na základě provedené rešerše si pro svoji práci vybral metodu Mask R-CNN, pro jejíž implementaci použil knihovnu detectron2. Metodu experimentálně ověřil na několika dodaných novinových datasetech, které si ještě musel předzpracovat. Provedené experimenty zpracoval a kriticky zhodnotil. Malou výtku mám k rozsahu kapitoly 2 popisující stávající metody detekce objektů. Popis jednotlivých metod je příliš stručný, což je na úkor srozumitelnosti.	

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>	
Odbornou úroveň hodnotím kladně.	

Formální a jazyková úroveň, rozsah práce	B - velmi dobře
<i>Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.</i>	
Po jazykové stránce nemám výhrad. Práce je napsána velice srozumitelně, s malým množstvím chyb a překlepů. Pozitivně hodnotím i doprovodné obrázky, které napomáhají pochopení problematiky. Kapitola 2 je zbytečně vertikálně členěna do hloubky dva, protože v hloubce jedna je pouze jedna podkapitola. V textu jsou často uvedeny neúplné odkazy ve tvaru x.y namísto Section x.y nebo Figure x.y.	

Výběr zdrojů, korektnost citací	A - výborně
<i>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.</i>	
Počet i výběr citovaných zdrojů je dostatečný.	

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.

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.

Student prokázal schopnost proniknout do daného tématu. Experimenty jsou dobře zpracovány a odprezentovány.

Do diskuze mám následující otázky:

1. Z tabulek 5.5, 5.7 a 5.8 se zdá, že čas potřebný pro natrénování modelů příliš nekoreluje s velikostí trénovací množiny. Čím to je?
2. Dále, ten výpočetní čas je v nižších jednotkách hodin. Není možné, že by se úspěšnost modelů zvýšila, kdybyste nechal algoritmus běžet po více než 1500 iterací?

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

Datum: 31.5.2021

Podpis: