prof. Ing. Boris Šimák, CSc.
ČVUT v Praze, FEL
Technická 2
166 27 Praha 6
tel.: (+420) 224 35 2203
e-mail: simak@fel.cvut.cz

Oponentský posudek disertační práce

Doktorand:

Ing. Pavol Harár

Název práce:

AUDIO CLASSIFICATION WITH DEEP LEARNING ON LIMITED DATA SETS

Předložená dizertační práce odpovídá oboru disertace Teleinformatika. Řeší velmi aktuální problematiku. Lze jí považovat za průkopnickou práci při zparacování problematiky zejména patologicky postižených hlasivek mluvčího. Výzkum směřuje do oblasti odstranění nevýhod subjektivních metod diagnostiky hlasu. Hlavním cílem práce je výzkum v oblasti nových přístupů prediktivního modelování založeného na technice hlubokého učení (DL) s využitím omezené sady zvukových dat.

Práce je napsána v angličtině, představuje soubor článků, doplněných průvodním textem, který ilustruje motivaci zaměření výzkumu a publikačních příspěvků v souvislosti s vlastním výzkumem autora.

Je rozdělena na tři klíčové části:

- Preambule
- Publikace
- Appendix.

V závěru je uvedeno CV autora.

V úvodní prvé části Preambule je popsána stručně oblast výzkumu. Je zaměřena na:

Hluboké učení;

Zpracování digitálních audiosignálů;

Automatickou analýzu lékařských zvukových dat.

Nakonec této kapitoly se disertant věnuje retrospektivně cíli práce, výzkumu nových přístupů DL založeného na prediktivním modelování na základě omezené množiny audio dat se zaměřením na hodnocení patologických anomálií v hlase. Cíl rozdělil dizertant na pět dílčích oblastí.

Cíl 1: *Prozkoumat specifiku analýzy lekařských hlasových dat pomocí hlubokého učení DL.*

Zde autor představuje experimentální přístup s přirozeným zánamem hlasu při hledání end-to-end systémů detekce patologie hlasu, který by mapoval prvotní záznam na odpovídající výstupy při

použití metod DL. Takové experimenty by také měly ukázat konkrétní povahu vlastností hlasových vzorků a způsob, jak s nimi nakládat při určování charakteristik získaných na základě DL.

Cíl 2: *Identifikovat budoucí architektury hlubokých neuronových sítí DNN (Deep Neural Network) s ohledem na systémy AVCA (automatický systém pro analýzu stavu hlasu).*

Úlohou bylo otestovat klíčové stavební bloky DNN používané v počítačovém vidění a analýza časových řad, jmenovitě v konvolučních neuronových sítích CNN a dlouhe vrstvy krátkodobé paměti (long short – term memory layers – LSTM) s očekávanou automatickou extrakcí vlastností.

Cíl 3: *Přezkoumat dostupné zdroje dat a jejich omezení.*

Disertant se plánoval zaměřit na přezkoumání dostupných použitelných zdrojů dat, určení, jaké je rozdělení zdravých vs. dysfonických vzorků a jaké je rozdělení (distribuce) zaznamenaných typů patologie vzorků a účel kombinace databází.

Cíl 4: *Objasnit, jaké vstupní a cílové reprezentace jsou užitečné*

Konkrétně trénovat modely používající přirozené vlnové průběhy a standardní časovo-kmitočtové reprezentace a porovnávat výsledky s ručně nalezenými řečovými vlastnostmi. Kromě toho určit, jaké další získané parametry, například jako je pohlaví, věk, stupeň dysfonie atd. ovlivňují schopnosti modelování a navrhnout možnosti předefinování úkolu změnou cílů.

Cíl 5: *Navrhnout protiopatření k vysoké poptávce po velkých souborech dat*

Zájem autora bylo nalézt postupy, které budou profitovat z trénování na omezeném souboru dat.

Druhá kapitola reprezentuje souhrn vybraných publikací autora, vztahujících se k tématu disertace. Příspěvky obsahem odpovídají zdrojovým publikacím a mají zde, v rámci dizertační práce, fixní strukturu. Obsah odpovídá původním publikacím. Názvy příspěvků jsou následující.

I *Voice Pathology Detection Using Deep Learning;*

II *Towards Robust Voice Pathology Detection;*

III *On Orthogonal Projections for Dimension Reduction and Applications in Augmented Target Loss Functions for Learning Problems;*

IV *Gabor Frames and Deep Scattering Networks in Audio Processing;*

V *Improving Machine Hearing on Limited Data Sets.*

Třetí část práce je věnována závěrečné diskusi nad dosaženými výsledky a směru dalšího výzkumu. Výsledky v rámci disertační práce, získané při řešení konkrétních výzkumných úkolů, jsou v rámci finální diskuse spojeny s jednotlivými vytčenými cíli práce.

V rámci přílohy je uvedeno CV autora, dávající náhled nad jeho odbornou kariérou a publikační aktivitou.

Disertant se zabýval v rámci spolupráce ve výzkumnémo týmu především vytvořením experimentálních numerických výsledků a jejich vizualizací. Příspěvky byly podpořeny řadou projektů, jak v rámci výzkumného centra SIXT při VUT v Brně, tak i na výzkumných zahraničních pracovištích na Universitě v Las Palmas de Grand Canaria a na Univerzitě ve Vídni.

Vlastní přínos autora spatřuji v návrhu a ověření nových přístupů v oblasti nástrojů detekce patologických nálezů z hlasu pacientů, kde prezentované publikace v dizertační práci jsou výstupem jeho dlouhodobé výzkumné aktivity. Výběr publikací v časopisech s IF dobře pokrývá prezentované cíle disertace.

Příspěvek *Voice Pathology Detection Using Deep Learning* je v souladu s cíli 1 a 2. Byly prezentovány důležité zkušenosti jak z hlediska testovaných přístupů DL s konvoluční neuronovou sítí

CNN, disertant se zaměřil i na úspory v oblasti potřebného rozsahu trénovacích dat pro DL s výhodou automatické extrakce příznaků.

Příspěvek *Towards Robust Voice Pathology Detection* a jeho závěry odpovídají záměrům cíle 3.

Cenné jsou praktické zkušenosti s kombinací několika databází s dysfonickými vzorky z hlediska jejich slučitelnosti v rozsáhlejší databázi. Problémem, který doktorand ověřil, je, že databáse nejsou konzistentní, jsou vytvářeny v různých jazycích a neobsahují identické promluvy. Vzorky s patologickými hlasy jsou nerovnoměrně rozloženy v databázi. Využitelné jsou pouze podmnožiny pečlivě vybraných vzorků. Proto experimenty byly provedeny na samohlásce /a/ s využitím několika databází a tím byl získán dostatečně velký soubor dat. Zkušenosti dávají možnost vytvořit matematický model pro objektivní detekci patologických nálezů.

Na základě výsledků experimentů diskutovaných v tomto příspěvku, použitím různých vstupních dat jako přirozeného hlasu, spektogramu, mel kmitočtových kepstrálních koeficientů (Mel-frequency cepstral coefficients - MFCC), konvenčních dysfonických vlastností, jejich kombinací a s využitím klasifikátorů k extrakci vlastností hlasu je splněn cíl čtvrtý.

Publikace III – V pokrývají pátý cíl disertační práce zaměřený na nalezení vhodné transformace vstupních a cílových dat a získat tak dostatečnou množinu relevantních dat s cílem jejich omezení pro praktické testování na bázi hlubokých neuronových sítí DNN. V numerických experimentech je představeno zlepšení výkonu konvolučních neuronových sítí trénovaných na omezené množině zvukových dat a navržených časově-frekvenčních reprezentacích "Gabor" a "Mel" rozptylech.

Článek *On Orthogonal Projections for Dimension Reduction and Applications in Augmented Target Loss Functions for Learning Problems* se věnuje vazbě mezi dvěma předměty při redukci dimenze a zabývá se aplikováním ortogonální projekce na úlohu hlubokého učení a zavedl obecnou funkci rozšířené cílové ztrátové funkce. Redukce dimenze je řešena ortogonální projekcí. Výsledky výzkumu disertant aplikoval na segmentaci klinického obrázku a klasifikaci hudební informace. Navržená funkce zlepšuje přesnost.

V článku *Gabor Frames and Deep Scattering Networks in Audio Processing* autor zavedl extractor vlastností, založený na Gaborovových rámcích a Mallatově rozptylové transformaci - Gaborův rozptyl. Implementoval Gaborův rozptyl do knihovny v jazyku Python. Využití Gaborovy rozptylové transformace poskytuje lepší výsledky ve srovnání s užitou Gaborovou transformací, zejména je-li omezené množství vzorků pro trénování.

V článku *Improving Machine Hearing on Limited Data Sets* je porovnán standardní vstupní mel-spektrogram s nově navrženým mel rozptylem. Pro ověření výsledků je k dispozici odkaz na repozitář s uloženým pogramem.

U každého příspěvku doktorand vymezuje svůj podíl na vytvořené publikaci. Vlastní přínos disertanta spatřuji zejména mimo experimentálních výsledků i v oblasti návrhu nových časově-frekvenčních reprezentací a jejich využití při diagnostice patologických nálezů pacientů. Za významné výsledky v teoretické oblasti považuji příspěvky IV a V.

Publikace jsou často citovány. Hirschův index má doktorand 4 dle Google Schollar a 3 podle Scopusu.

Lze konstatovat, že cíle dizertační práce byly splněny.

Předložená disertační práce je kvalitní a to jak po stránce obsahu, tak i zpracování. Text disertace je napsán srozumitelně, grafické zpracování je na dobré úrovni. Jádro dizertační práce bylo dostatečně publikováno. V práci je naznačen směr dalšího výzkumu na základě získaných teoretetických i experimentálních výsledků.

Autor práce prokázal svoji odbornou erudici, osvojil si dobře zásady a metody vědecké práce, včetně týmové spolupráce. Je uznávaným odborníkem v dané oblasti, absolvoval řady vyzvaných přednášek.

K formálnímu zpracování práce mám následující připomínky:

- V práci jsou navržené postupy a výsledky popsány stručně.
- Práce obsahuje minimum překlepů.
- Velký rozsah je věnován popisu experimentálních přístupů ve srovnání s teoretickým přínosem publikací.
- Chybí seznam zkratek. Pojmy jsou sice vysvětleny či označeny v textu, ale tento seznam by byl přínosný. Některé zkratky jsou použity dříve, než jsou definovány.

Pro diskusi navrhuji následující otázky:

- Jak budou dále rozvíjeny navržené postupy z hlediska možnosti nasazení v praxi?
- Co očekáváte od přínosu dalších vrstev DL modelu, jak zajistíte, aby nedošlo k přetrénování?
- Které výzkumné týmy se zabývají obdobnou problemetikou v ČR?
- Jak jsou chráněny dosažené výsledky?

Doktorand prokázal schopnost samostatné vědecké práce. Disertační práce splňuje podmínky uvedené v § 47 odst. 4 zákona č. 111/1998 Sb., o vysokých školách. S ohledem na výše uvedené skutečnosti hodnotím disertační práci Ing. Pavola Harára jako úspěšnou a doporučuji ji k obhajobě a po úspěšné obhajobě doporučuji udělení titulu Ph.D.

prof. Ing. Boris Šimák, CSc.

ČVUT v Praze, FEL

V Praze dne 6. 10. 2019