

Oponentský posudek

na doktorskou disertační práci

Ing. Karla Petránka

Detekce objektů ve videosekvenci

pro obor Aplikovaná informatika na Univerzitě Hradec Králové

školitelka prof. RNDr. Eva Milková, PhD.

Oponent: Doc. Ing. Arnošt Veselý, CSc.

Česká zemědělská univerzita, Provozně ekonomická fakulta,

Katedra informačního inženýrství

Předložená práce je z oblasti hlubokého učení (deep learning). Tato oblast se v posledních letech velmi rychle rozvíjí a výsledky, které byly v této oblasti dosaženy, umožnily podstatný pokrok v oboru umělé inteligence. Metody hlubokého učení dovolují efektivní rozpoznávání objektů na fotografických snímcích a videu, a mají důležitý dopad v praktických aplikacích umělé inteligence. Problematika, kterou se autor zabývá je tedy velmi aktuální a velmi dobře aplikovatelná v praxi.

Cíle práce

Autor si pro svou práci stanovil následující cíle:

- 1) Navrhnout a realizovat unikátní datovou sadu pro testování algoritmů rozpoznávání objektů v obrazech a videosekvencích
- 2) Vylepšení způsobu extrakce atributů
- 3) Zlepšení detekce umístění objektu
- 4) Sledování objektu ve videosekvenci

Pokud se týče prvního cíle, je snadné si představit jak postupovat, ale vlastní realizace datové sady, která musí mít pro deep learning řádově desetitisíce prvků je namáhavá.

Problematika, která souvisí s druhým cílem, tj. extrakce atributů je důležitá pro klasické metody rozpoznávání, které nepoužívaly konvoluční sítě. Konvoluční sítě extrakci atributů nepotřebují, protože provádějí extrakci atributů samy automaticky.

Na řešení třetího a čtvrtého cíle je zaměřen současný výzkum ve světě a v této oblasti deep learningu se intenzívně pracuje.

Struktura předkládané disertační práce je následující:

Kapitola 1 je úvodem a v kapitole 2 jsou formulovány 4 cíle, které autor chce ve své práci řešit.

Kapitola 3 obsahuje analýzu současného stavu v oblasti detekce objektů. Autor nejdříve popisuje principy algoritmů pro extrakci atributů. Následuje velmi stručný úvod do klasifikačních algoritmů. V dalším paragrafu autor představuje problematiku a základní pojmy z oblasti neuronových sítí. Výklad je veden z hlediska deep learningu, tj. vysvětluje pojmy, které se v klasické teorii neuronových sítí

nevyskytovaly, jako je například batch normalizace, dropout, neurony s rektifikovanou lineární výstupní funkcí apod. V tomto paragrafu autor také stručně popisuje architektury konvolučních sítí, které se pro detekci neuronových sítí v současné době používají. Poslední paragraf třetí kapitoly popisuje principy detekce a sledování objektů ve videosekvenci, včetně úspěšných algoritmů publikovaných v posledních několika letech. V posledním paragrafu autor popisuje datové sady, které se pro detekci objektů v obraze nebo videosekvenci v současné době používají.

Lze říci, že autor na velmi omezeném prostoru asi 30 stránek podal přehledný popis oblasti, jejíž problematiku ve své práci řeší. Výklad je přehledný. Je sice neformální, ale formulace jsou přesné alespoň do té míry, jak to neformální jazyk dovoluje.

Oblast, které se výklad týká, je tak široká, že nebylo možné všechny důležité věci do výkladu zařadit. Názor na to, které věci bylo třeba zařadit a které bylo možné vynechat je jistě velmi subjektivní. Z mého hlediska se příliš nepovedl výběr látky v paragrafu o klasifikačních algoritmech. Jedná se o dosti rozsáhlou problematiku. Autor zde uvádí logistickou regresi, rozhodovací stromy a boosting, což těžko lze považovat za reprezentativní výběr klasifikačních algoritmů. Kromě toho v dalším textu se tyto klasifikační algoritmy ani nepoužívají.

V paragrafu o neuronových sítích mi chybí stručný obecný výklad o konvolučních sítích a problematice jejich učení. Na konvolučních sítích je celá problematika deep learningu založena a s konvolučními sítěmi se autor ve své práci zabývá neustále.

Výsledky práce autora jsou předmětem kapitoly 4. V prvním paragrafu čtvrté kapitoly je popsána datová sada pro detekci a sledování objektu ve videosekvenci, kterou autor navrhl a realizoval. Autor se ujal dosti pracného úkolu, většinou se pro testování algoritmů používají sady dostupné na internetu. Zde je škoda, že autor podrobněji nepopsal, jakými softwarovými prostředky sadu vytvořil a jaké přitom získal zkušenosti. Bylo by to dobrým vodítkem pro ostatní, kteří o vytvoření své vlastní datové sady rovněž uvažují.

V dalším paragrafu 4.2 autor popisuje výsledky, které získal při řešení svého druhého cíle, a sice zlepšení extrakce atributů. Autor navrhl úpravu algoritmu LBP (Local Binary Patterns) a takto upravený algoritmus předřadil jednak před klasifikační standardní hustou sítí a jednak před konvoluční sítí. Experimenty autor prováděl na standardně používaných datových sadách MNIST (rukou psané číslice) a CIFAR10 (pixelové barevné obrázky deseti kategorií). Z uvedených výsledků vyplývá, že jím upravený algoritmus dosahuje většinou větší přesnosti klasifikace a je časově méně náročný. Autor ale sám připouští, že LBP transformaci atributů není vhodné předřazovat před konvoluční sítí. Pokud je obraz vstup vložen přímo na vstup konvoluční sítě, je výsledná přesnost klasifikace výrazně lepší.

V paragrafu 4.3 autor popisuje metodu, kterou navrhnul pro sledování objektů ve videosekvenci. Pro detekci pohybujícího se objektu autor použil upravenou architekturu konvoluční sítě SharpMask ve spojení s algoritmem založeným na částicovém filtru. Svoji metodu ověřil a porovnal ji s výsledky, které získal na stejné datové sadě s několika k tomuto účelu standardně používanými algoritmy, jako jsou například YOLOv2, GOTURN atd. Z příložené tabulky je vidět, že navržená metoda dávala o něco lepší výsledky než nejuspěšnější z nich metoda YOLOv2.

V posledním paragrafu 4.4 autor využil upravenou architekturu SharpMask natrénovanou na datové sadě COCO pro segmentaci objektů na datové sadě DAVIS. Její úspěšnost se pak posuzuje autorem definovaným F skórem. Z příložených obrázků se segmentace jeví jako úspěšná.

Paragrafy 4.3 a 4.4 považuji za nejdůležitější část dizertační práce. Autorem navržený způsob sledování objektů ve videosekvenci je originální a dává velmi dobré výsledky.

K paragrafům 4.3 a 4.4 mám následující výhradu: Vzhledem k tomu, že tato část již není rešerší, ale jádrem celé práce, autor měl tuto část zpracovat podrobněji a pro přesnou formulaci použít matematický formalismus. Autor měl přesně definovat částicový filtr a jeho vlastní algoritmus, který je na modifikaci částicového filtru založen. Rovněž měla být přesně popsána architektura použité konvoluční sítě. Obr. 20 je pouze hrubým popisem její architektury. Není například patrné, kde v ní byla použita batch normalizace, kde byl použit dropout a jakou měl velikost, jaké byly kroky u konvolučních a maxpooling vrstev atd. Rovněž parametry týkající se učení sítí nejsou popsány.

V poslední páté kapitole autor shrnuje získané výsledky a konstatuje naplnění cílů práce. S tímto konstatováním souhlasím.

Celkové hodnocení práce:

Práce se zabývá aktuální problematikou se značným dopadem v praktických aplikacích umělé inteligence.

V práci se autor odvolává na celou řadu prací zahraničních autorů, většinou z posledních deseti let. To svědčí o tom, že se autor důkladně seznámil se současným stavem řešené problematiky a pracuje na aktuálních problémech.

Výsledky své práce autor publikoval v odborném tisku. Je autorem nebo spoluautorem celé řady článků, z nichž nejméně 4 se zabývají problematikou řešenou v disertační práci.

Práce je dobře napsána a je srozumitelná, i když nutně používá celou řadu oborově specifických pojmů a zkratk.

K disertační práci mám následující konkrétní výhrady a připomínky:

Své výhrady k práci jsem uvedl výše.

Navrhuji, aby doktorand při své obhajobě podal

- 1) formálnější popis navržené metody pro sledování objektu ve videosekvenci a také přesnější popis konvoluční sítě, která je její součástí
- 2) podrobnější popis výsledků, které získal při jejím ověřování

Závěr posudku:

Práce se zabývá problematikou, která je velmi aktuální a pro praktické aplikace umělé inteligence velmi důležitá. Autor prokázal, že se s danou problematikou do hloubky seznámil a že se v ní velmi dobře orientuje. Navrhnul originální efektivní algoritmus pro sledování objektů v obrazech a videosekvencích a jeho vlastnosti ověřil na vlastní datové sadě.

Konstatuji, že disertační práce prokazuje předpoklady autora k samostatné tvořivé vědecké práci.

Disertační práci doporučuji k obhajobě a po jejím úspěšném absolvování doporučuji udělení vědecké hodnosti PhD v oboru Aplikovaná informatika.

V Praze dne 7.2.2018

Doc. Ing. Arnošt Veselý, CSc.