

Oponentský posudek na diplomovou práci

Autor: David Novotný
Titul: Large scale object detection

Vypracoval: RNDr. Barbara Zitová, PhD.
Ústav teorie informace a automatizace AV ČR, v.v.i.,
Pod Vodárenskou věží 4, Praha 8

Předložená diplomová práce se zabývá problematikou klasifikace a detekce objektů scény náležejících do mnoha tříd. Autor vycházel z dvou uznávaných přístupů k uvedenému problému – Fischer kernel funkcí a konvolučních neuronových sítí (CNN). Nově předložená metoda, použitelná jak pro klasifikaci, tak pro detekci objektů z mnoha tříd, je motivována oběma směry. V druhé fázi práce se autor věnoval paměťové náročnosti navrženého postupu a jejího řešení pomocí selekce a komprese příznaků. V poslední fázi se autor zabýval testováním navržených schémat zpracování obrazových dat a porovnání dosažených výsledků s referenčními hodnotami pro uvedené klasifikační a detekční třídy úloh.

Na vstupu metody se předpokládají obrazová data objektů pro klasifikaci, případně detekci, kde se žádá navíc určení polohy objektu. Autor pracuje s datasetem Pascal VOC 2007 (viz. citace v práci). Volba testovacích dat je vhodná pro dané téma práce, jedná se o veřejně dostupný dataset s anotacemi zahrnutých 20 tříd objektů.

V zadání práce byla úloha formulovaná následovně – student se měl seznámit se současnými metodami pro klasifikaci objektů do tříd, zvolenou metodu naimplementovat a prostudovat/vyhodnotit a navrhnout vylepšení a toto také zhodnotit. Uvedená úloha klasifikace/detekce objektů s možností přináležení do mnoha tříd je rozhodně aktuální úloha, která s rostoucím množstvím obrazových dat a snahou o jejich automatické porozumění nabývá na důležitosti. Navrhované řešení vychází ze současných uznávaných metod (CNN, klasifikace pomocí *kernel* funkcí).

Student se seznámil s existující literaturou v oblasti klasifikace objektů z mnoha tříd, stejně tak nastudoval potřebnou teorii k jednotlivých přístupům a jejich efektivního vyhodnocování. Seznam literatury zahrnuje aktuální výsledky v oboru. Získané poznatky shrnul a úspěšně prezentoval ve své práci.

Vlastní přínos studenta spočívá v návrhu fúze dvou standardních přístupů a jejich otestování a porovnání se současnou špičkou. Autor vyšel z teorie Fisher vektorů, které umožňují použít lineární klasifikátory pro klasifikaci objektů. Běžný postup, kdy je využíváno pro jejich výpočet modelování Gaussovských směsí, autor práce nahradil úvahou využít konvoluční neuronové sítě (softmax funkce). Navrhovaný postup místo *log-likelihood* funkce používané v standardní teorii pracuje s navrženou tzv. *pseudo-loglikelihood* funkcí, která vychází z CNN. Dále pak autor pracuje s takto odvozenými *Fisher kernel based* příznaky. Autor provedl diskuzi nástrah uvedeného postupu, přiznává, že místy chybí teoretické zdůvodnění a je nahrazeno empirickým pozorováním. Autor zpracoval 4 přístupy pro redukci počtu příznaků a jejich případnou kompresi (sekce 3.3). Navrhl dvě schémata kombinace získaných klasifikátorů, lišících se v okamžiku, kdy jsou data zkombinována. Pro porovnání efektivit navržených postupů a jejich srovnání autor provedl řadu experimentů, kdy jsou metody srovnány navzájem a také vůči referenčnímu postupu (pro klasifikaci a detekci zvlášť). Pro experimenty byly použity dostupné CNN architektury a modifikované algoritmy pro redukci počtu příznaků a pro klasifikátory. V závěru práce je uvedena diskuze a návrh

možného pokračování. Výkonnost nově navržených metod podle uvedených experimentů je většinou srovnatelná nebo lepší než srovnávané referenční postupy.

K diplomové práci mám několik dotazů a komentářů. Jeden se týká zadání práce – tam je důraz kladen na detekci objektů z mnoha tříd (jak text, tak doporučená literatura), práce působí spíše dojmem, že se autor věnoval úloze klasifikace. Z jakých důvodů byl původní směr upozaděn?

K jednotlivým částem práce pak mám následující:

- Pracujeme-li s jedním obrázkem ($n=1$) a váhami, biasem a aktivacemi, existuje intuitivní náhled na U_x (3.11)?
- Jaká je citlivost odhadu I na trénovací množinu?
- U brzké fúze (3.29) - neprojevuje se vliv rozdílného dynamického rozsahu hodnot pro příznaky CNN – FK a aktivit neuronu ?
- Obr. 4.1 – předpokládám, že popis neodpovídá plně obsahu obrázku.
- Od podání práce, podařilo se ověřit experimenty na TUNE RNK síti, zmíněné v textu?

Z formálního hlediska je práce dobrá, poskytuje dostatečný úvod do problematiky a popis navržených postupů. Ve značení by se dalo ještě dosáhnout zlepšení (chybí označení (2.5) jakožto Fisher vektoru, přesnější zavedení *log-likelihood* funkce). V textu jsou nějaké překlapy/nejasnosti, ale asi nejvíc bijící do očí je opakované použití formulace likelihood. Práce je napsaná srozumitelnou angličtinou, ilustrace jsou voleny vhodně a názorně.

Celkově působí práce velmi dobře. Představený přístup představuje řešení, které experimentálně vyčerpávajícím způsobem prokázalo svou použitelnost, zároveň je v textu snaha vše teoreticky obhájit. Autor navrhl jak nový postup, tak řešil následně jeho paměťovou náročnost a vlastní implementovatelnost. Experimentálně srovnal jednotlivé schémata a porovnal s existujícími referenčními postupy. Navíc podle toho, co jsem našla na webu (<http://www.vlfeat.org>) se autor v oblasti pohybuje již delší dobu.

Doporučuji tuto diplomovou práci klasifikovat známkou A - „výborně“.

V Praze, 14.12. 2014

RNDr. Barbara Zitová, Ph.D.