Detekce klíčových bodů pomocí konvoluční neuronové sítě

Ivan Gruber[1]

1 Úvod

Úloha detekce klíčových (významných) bodů na lidské tváři je v dnešní době rozšířenou úlohou. I přes veškerou snahu však není plně vyřešena a to především kvůli tomu, že se vzhledem k rozmanitým externím (osvětlení, póza, okluze) a interním (výraz tváře, stárnutí) podmínkám jedná o velmi komplexní a složitou úlohu.

Existuje mnoho rozličných algoritmů, které detekci klíčových bodů řeší. Tyto algoritmy lze rozdělit podle jejich přístupu k řešení do tří skupin: top-down a bottom-up a jejich kombinace. Mezi typického představitele top-down přístupu můžeme zařadit Active Appearance Model (AAM) (viz Cootes a Taylor (2004)). Do druhé skupiny lze zařadit především metody založené na neuronových sítích (v posledním roce se tímto přístupem zabýval například Zhang a spol. (2015)). Tímto přístupem se bude dále zabývat i tato práce.

2 Dataset a Augmentovaná data

Pro trénování neuronové sítě jsem zvolil databázi Helen (více viz Vuong a spol. (2012)). Databáze byla vytvořena z obrázků z Flickeru, na každém snímku bylo manuálně anotováno 68 klíčových bodů. Databáze obsahuje 2330 RGB snímků, které jsem dále rozdělil na trénovací (obsahující 2000 snímků) a development (obsahující 330 snímků) sadu.

V prvním kroku přípravy dat bylo třeba na každém snímku detekovat danou tvář. Použil jsem dva přístupy - pomocí Viola-Jones detektoru (viz Viola a Jones (2004)) a ořez vzhledem k anotovaným klíčovým bodům. Druhý přístup samozřejmě na reálná data nelze využít, nicméně zde byl použit k vytvoření vyššího počtu dat. Velikost oblasti zájmu (ROI) byla unifikována na 120×100 pixelů.

Takto oříznuté ROI byly augmentovány několika způsoby: převrácení (flip), rotace (v kladném i záporném směru otáčení), translace, transformace jasu (pomocí aditivního Gausse), rozmazání, zašumění klíčových bodů. Ve finále testovací sada obsahovala 54958 snímků a development sada 9663 snímků.

Konvoluční neuronová síť byla implementována ve frameworku Caffe. Síť byla optimalizována na základě Euklidovy vzdálenosti nalezených a anotovaných bodů pomocí metody stochastic gradient descent (SGD). Během experimentů jsem otestoval několik architektur sítě.

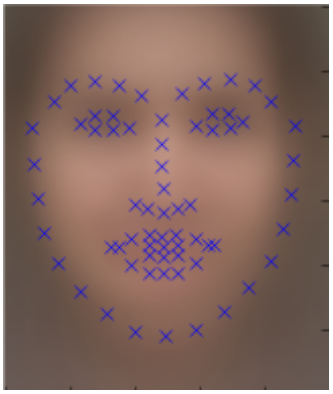
3 Experimenty a výsledky

V prvním experimentu byla síť natrénována v celkem 50000 iteracích, přičemž na konci dosáhla chyby přibližně 4.4 pixely na jeden bod na development datech. V druhém experimentu jsem nejprve spočetl průměrné pozice klíčových bodů přes celou trénovací množinu

[1] student doktorského studijního programu Aplikované vědy a informatika, obor Kybernetika,
e-mail: grubiv@ntis.zcu.cz

(viz Obrázek 1). Dále jsem u každého snímku vypočetl deltu tohoto průměru a anotovaných bodů. Síť byla optimalizována (opět v 50000 iteracích) na základě Euklidovy vzdálenosti takto vypočtených rozdílů. Chyba byla snížena na 3.5 pixelu na bod na development datech.

Tyto prvotní experimenty ukázaly významné snížení chyby při trénování rozdílů místo absolutních souřadnic bodů. Síť lépe reagovala na extrémní případy natočení tváře a taktéž se dokázala lépe naučit závislosti mezi jednotlivými body (jejich strukturu). Celková chyba je sice stále signifikantní, nicméně hlavním přínosem těchto experimentů nebylo nalezení optimálního řešení, nýbrž srovnání dvou výše uvedených přístupů k trénování. V budoucím experimentu bych rád natrénoval neuronovou síť pomocí rozdílových obrazů (získaných odečtením průměrné tváře od daného snímku).



Obrázek 1: Průměrná tvář s průměrnými klíčovými body.

4 Závěr

Neuronové sítě jsou všestranným nástrojem, dostatečně rychlým a přesným pro řešení problému detekce klíčových bodů. Výše uvedené experimenty poslouží jako výchozí bod pro moji další práci s neuronovými sítěmi.

Poděkování

Příspěvek byl podpořen grantem Západočeské Univerzity, číslo projektu SGS-2016-039. Dále bych rád poděkoval za přístup k vypočetním a úložným zařízením vlastněných společností National Grid Infrastructure Metacentrum, poskytovaných v rámci projektu CESNET LM2015042.

Literatura

Cootes, T.F., Taylor, C. J., 2004. *Statistical Models of Appearance for Computer Vision*. Imaging Science and Biomedical Engineering, University of Manchester, pp. 149–163.

Zhang, J. , Kan, M., Shan, S., Chen, X., 2015. *Leveraging Datasets with Varying Annotations for Face Alignment via Deep Regression Network*. In: Proceedings of IEEE International Conference on Computer Vision (ICCV), pp. 3801–3809.

Viola, P., Jones, M., 2004. *Robust Real-time Face Detection*. International Journal of Computer Vision, vol.57(2), pp. 137–157.

Vuong, L., Brandt, J., Zhe, L., Boudev, L., Huang, T., 2012. *Interactive Facial Feature Localization*. In: Proceedings of ECCV2012, pp. 679–692.