



Detekce významných bodů na lidské tváři pomocí neuronové sítě

Ivan Gruber¹

1 Úvod

Detekce významných bodů je často velmi složitý a náročný úkol. Obzvláště detekce obličejových rysů se ukázala jako nesnadná a komplikovaná úloha, jelikož se velmi liší nejen u jednotlivých osob, ale často i pro jediného člověka, a to díky velkému množství variací 3D pozice (natočení, vzdálenost), světelných podmínek nebo výrazu tváře.

I přes všechny výše zmiňované problémy se v posledních letech jedná o vyhledávanou úlohu, jelikož predikce pozice těchto bodů má široké využití: sledování tváře ve videu, analýza výrazu obličeje, odhalování dysmorfických známek v obličeji pro lékařské účely a nebo například biometrie.

Obvykle je přístupováno k detekci dvěma rozličnými způsoby. Prvním způsobem je vyhledávání pomocí Support Vector Regresoru (SVR) trénovaného na predikci pozice bodů z vhodného obrazového deskriptoru (obvykle založeného na LBP histogramech), viz současná práce pána Taigmana a spol. (2014). Druhým přístupem je natrénování vhodné neuronové sítě a detekce bodů s její pomocí. Tato práce dále popisuje pouze detekci bodů pomocí neuronové sítě.

2 Použitá databáze a neuronová síť

Pro trénování neuronové sítě byla využita databáze MUCT, více viz Milborrow a spol. (2008). Databáze obsahuje celkem 3755 RGB snímků s rozlišením 640x480 pixelů, které bylo pro účely trénování neuronové sítě sníženo na pětinu (tedy na 128x96). V databázi je zachyceno celkem 276 různých osob všech barev pleti, rozličného věku a obou pohlaví. Každá osoba byla vyfocena simultánně z pěti kamer, přičemž každá fotila dotyčný subjekt z jiného úhlu. Dále byla většina osob stejným způsobem zachycena ve třech různých světelných podmínkách.

Na každém snímku výše uvedeném databáze bylo anotováno 76 významných bodů lidského obličeje (viz Obrázek 1). Pokud na nějaké fotce daný bod nebyl viditelný, byla mu přidělena souřadnice 0,0 (levý horní roh obrázku). Tyto body byly vynechány z pozdějšího trénování neuronové sítě. Po této redukci nám zbylo 70 bodů pro každý obrázek. Všechny obrázky byly dále rozděleny na dvě množiny, trénovací a development sadu, a to v poměru 3300 a 475 obrázků.

Pro samotnou neuronovou síť byl zvolen framework Caffé, více viz práce Yangqinga a spol. (2014). Trénování sítě bylo optimalizováno pomocí euklidovy vzdálenosti nalezených a anotovaných bodů. Během testování bylo vyzkoušeno několik architektur sítě, přičemž nejlepší výsledky poskytovala architektura skládající se ze čtyř vrstev. L1 obsahovala konvoluční vrstvu, ReLU a maxpooling, L2 a L3 se skládaly z dalších konvolučních a ReLU vrstev, a poslední L4 se skládala z jedné fully-connected a ReLU vrstvy.

¹ student doktorského studijního programu Aplikované vědy a informatika, obor Kybernetika, e-mail: Ivan.Gruber@seznam.cz



Obrázek 1: Snímek z MUCT databáze se 70 anotovanými body

3 Výsledky a využití

Výsledná síť byla natrénována v celkem 20000 iteracích, přičemž na konci dosahovala průměrné chyby 2 pixely na jeden bod na development datech. Detekce bodů při čelním pohledu byla natrénována s minimální chybou, chyba při jiných pohledech byla větší. Problémy síti činily především osoby tmavší pleti, a to nejspíše z důvodu nedostatku snímků, na kterých by byly zachyceny. Výhodou celé sítě je především její rychlost, detekce bodů na jednom snímku trvá v průměru 0.03 sekundy.

V budoucnu je plánováno tyto významné body využít k vypočtení pozice a natočení neznámé lidské tváře pro následný fitting morfovateľného modelu lidské tváře. Vzhledem k tomuto plánovanému využití jsou získaná data poskytnutá natrénovanou neuronovou sítí naprosto postačující.

4 Závěr

Neuronová síť se prokázala být vhodným, rychlým a dostatečně přesným nástrojem pro detekci klíčových bodů na lidském obličejí. Výsledky z tohoto experimentu budou v budoucnu využity k dalším experimentům, a to především k experimentům s morfovateľným modelem lidské tváře.

Poděkování

Práce byla podpořena grantovým projektem SVK1-2015-015.

Literatura

- Milborrow, S., Morkel, J., Nicolls, F., 2008. *The MUCT Landmarked Face Database*. University of Cape Town, Cape Town.
- Taigman, Y., Ming, Y., Ranzato, M., Wolf, L., 2014. *DeepFace: Closing the Gap to Human-Level Performance in Face Verification*. Columbus, Ohio.
- Yangqing, J., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrrell, T., 2014. *Caffe: Convolution Architecture for Fast Feature Embedding*.