

Posudek oponenta bakalářské práce

České vysoké učení technické v Praze
Fakulta jaderná a fyzikálně inženýrská
Katedra matematiky

Autor práce: Thi Thu Hien Nguyenová
Název práce (EN): Convolutional neural network-based human body detection and tracking in the interior
Název práce (CZ): Detekce a sledování lidského těla ve vnitřním prostoru pomocí konvolučních neuronových sítí
Datum zadání: 31. 10. 2020
Datum odevzdání: 7. 7. 2021
Studijní obor: Aplikované matematicko-stochastické metody
Autor posudku: Ing. Adam Novozámský, Ph.D.
Role: Oponent
Pracoviště: Computer Vision Lab
Faculty of Informatics Institute of Visual Computing & Human-Centered Technology
TU Wien, Favoritenstr. 9/193-1, A-1040 Vienna, Austria

Ačkoli je odhad pozice a natočení hlavy problém, který se řeší v komunitě počítačového vidění poměrně dlouho, nelze říct, že by šlo o zcela vyřešenou věc. Proto studium dosavadních metod a návrh nových je něco co budeme potkávat jak v literatuře, tak studentských pracích i v následujících letech. Zadání bylo správně navrhnuté tak, aby si autorka mohla vyzkoušet všechny kroky pro úlohy hlubokého učení – tvorba a příprava dat, návrh sítě, trénování a evaluace výsledků a experimenty na reálných datech. Zároveň mohla tyto výsledky porovnat s klasickou metodou počítačového vidění.

K textu práce nemám zásadní výhrady. Text je logicky členěn do kapitol, tak jak by člověk očekával a při čtení textu nemá člověk pocit, že by se ztrácel. Vše je srozumitelně a jednoduše vysvětleno, jak na začátku práce v teorii sítí, při popisu metod, nebo na konci práce v experimentální části. Citace literatury i odkazy na jednotlivé části textu, obrázky a tabulky jsou uváděny, tak jak je zvykem.

Měl bych asi jen tyto stručné komentáře k textu:

1. Doporučil bych místo „vytrénování sítě“ používat spíše slovní spojení „natrénování sítě“, jak je standartně používáno v literatuře.
2. Pak jsem trochu zmaten v experimentální části, kdy autorka ukazuje úspěšnosti při trénování pro různé nastavení hyperparametrů. Je zde mluveno o různých verzích sítě, přičemž jsou měněny občas jen ty hyperparametry, které nemají vliv na samotnou architekturu modelu (počet epoch, batch-size, optimalizátor atd.). Bylo by tedy srozumitelnější, kdyby se mluvilo o verzi nastavení trénovacího procesu.
3. V práci se střídá jednotné a množné číslo při popisu, co bylo uděláno:
 - a. „Přišlo mi, že ...“
 - b. „Abychom byli schopni ohodnotit ...“Lépe by bylo zůstat u množného čísla v celém textu práce.
4. Na začátku experimentální části je uvedeno, že síť má jen 3 lineární vrstvy, mezi které jsou postupně přidávány aktivací a vypínací vrstvy. Pro takto malé sítě bych doporučil architekturu sítě vždy graficky znázornit určitě by to pro srozumitelnost pomohlo.

Výsledná navrhnutá síť je velmi jednoduchá a trénovací dataset není také dostačující (celkem 100 obrázků – po augmentaci 200), proto výsledky nedosahují přesnosti současných state-of-the-art metod používajících pro testování desítky tisíc různě natočených tváří. Nicméně si na práci cením, že autorka vytvořila vlastní dataset pomocí 3D Malování, vyzkoušela si augmentaci a síť, sice ne nijak hlubokou, si navrhla sama. Pokud tedy měla práce sloužit hlavně edukativně pro seznámení se s problematikou a jako příprava na práci diplomovou, splnila svůj účel. Zároveň je patrný velký kus odvedené práce jak při rešerši metod, tvorbě datasetu, tak v experimentální části.

V závěru experimentální části testovala autorka svoji metodu na snímcích z YouTube videa a trenažeru, kde jen demonstrovala spočtené úhly natočení. Škoda, že nepoužila některý ze zmiňovaných volně dostupných datasetů, kde je přístupné i ground-truth a mohla tak svoji metodu přímo porovnat s dostupnými metodami v literatuře.

Moje dotazy a podněty k práci:

1. V práci není uvedeno, jak velké byly snímky vytvořené v programu 3D malování.
2. Výsledná úspěšnost odhadu úhlů má poměrně vysokou chybovost (kolem 10°), domnívám se, že velkou roli na tom může mít samotné použití MTCNN na detekci bodů, kde posun jednotlivých bodů vůči sobě i o pixely může hrát velkou roli ve výsledné chybě. V práci to není uvedeno, ale nezkoušela jste tuto chybu spočítat, na tomto Vašem malém datasetu?
3. Nikde jsem v experimentální části, diskusi ani závěru nenašel zmínku o celkovém času potřebném na zpracování jednoho snímku o dané velikosti. Jen je uvedeno, že čas síťového řešení je srovnatelné s časem při geometrické metodě. Což není mimochodem příliš překvapující, když obě metody používají pro nalezení těch 5 hlavních bodů algoritmus MTCNN, který spotřebovává asi většinu výpočetního času. Máte tedy změřeno, jak dlouho trvá zpracovat jeden snímek o velikosti např. Full HD a jaké by bylo při tomto rozlišení FPS při použití metody na každý snímek bez využití pozice z předchozího framu?

I přes moje výše uvedené výhrady doporučuji bakalářskou práci vypracovanou slečnou Thi Thu Hien Nguyenovou klasifikovat výborně, tedy známkou **A**.

Místo, datum a podpis: Praha 5. 8. 2021

.....
Ing. Adam Novozámský, Ph.D.