

**Posudek na bakalářskou práci Adama Janicha
„Automatická analýza buněčných procesů ze snímků pořízených TIRF-SIM mikroskopem“**

vypracoval Jan Kamenický (oponent)

V bakalářské práci se student zabývá automatickým vyhodnocováním TIRF-SIM mikroskopických snímků endocytických procesů probíhajících na buněčné membráně živých buněk. Konkrétně je zde cílem nalézt endocytické váčky, které se v určitém stádiu jeví na snímcích jako jakési „kroužky“. Tyto struktury jsou ovšem pozorovatelné pouze v rozlišení, které je momentálně na hranicích technických (a fyzikálních) možností, data jsou výrazně zašuměna a po SIM rekonstrukci obsahují artefakty. Jedná se tedy o poměrně obtížnou úlohu a automatické zpracování takových dat může mikrobiologům opravdu pomoci.

V první kapitole je popsán úvod do biologie endocytózy a technologie snímání TIRF-SIM mikroskopie, a to velmi pěkně a čtivě. V druhé kapitole následuje stručný úvod do problematiky neuronových sítí, který je ovšem vzhledem k zaměření práce nedostatečný. Zcela chybí přehled běžně používaných sítí ať už obecně nebo konkrétně pro řešenou úlohu klasifikace (což je asi nejtypičtější úloha řešená konvolučními neuronovými sítěmi a článků na toto téma jsou tisíce). V literatuře není jediný článek zabývající se neuronovými sítěmi! Po technické kapitole 3 následuje příprava trénovacích dat. Kvalita a množství trénovacích dat je pro použití neuronových sítí klíčové, zmíněné počty zástupců jednotlivých kategorií vypadají dostatečně a dovedu si představit, že jejich příprava mohla být celkem pracná. V páté kapitole je popsána architektura použité sítě. Tady bych opět očekával více vzhledem k tomu, že to je klíčová část práce. Chybí jakákoliv diskuze nad tím, jakou architekturu zvolit tak, aby byla síť co nejlépe vhodná pro danou úlohu, resp. popis nějakého procesu, jak k dané architektuře student došel (inspirace z literatury, vlastní pokusy a omyly, apod.). V kapitole 6 je popsáno poměrně standardní načítání a augmentace dat před vstupem do sítě. Trénování a validace dat popsané v kapitole 7 jsou navrženy rozumně a ukazují dobrou úspěšnost.

Celá práce je dobře strukturovaná i slohově a gramaticky dobře napsaná, ovšem obsahuje mnoho překlepů a pravopisných chyb (což považuji v dnešní době kvalitní automatické kontroly pravopisu za dost zbytečné).

Další připomínky:

- Obr. 6.1. – zobrazeny jsou výřezy cca 16x16 z 20x20, což neodpovídá v textu popisovaným výřezům 20x20 z 32x32.
- Obr. 6.2. – zobrazené transformace (1., 2. a 4. ve spodním řádku) neodpovídají v textu popisovaným (v textu není zmíněna rotace ani transpozice podle vedlejší diagonály).

Dotazy:

- Zvolené ořezy (32x32 -> 20x20) zaručují zachování pouze středu vstupního obrázku 8x8. Je to dostatečné (zejména u dvojkroužku)?

- Při tvorbě trénovací množiny je zdůrazňována nutnost rovnoměrného zastoupení jednotlivých kategorií. Není naopak možné trénovat tak, jak jsou data získána a tím přirozeně naučit síť apriorním pravděpodobnostem jednotlivých kategorií v datech?
- Pokud je validační množina fixní pro všechny natrénované instance sítě, nemůže volba této množiny (byť jen mírně) ovlivnit výsledky? Neměla by se validační množina vybírat náhodně pro každou instanci zvlášť?

Mohu hodnotit pouze to, co je v textu práce, takže pokud student popsal vše, co zkoušel a implementoval, pak jsem nucen volit hodnocení „dobře“ (C). Pokud by se (např. v průběhu obhajoby) ukázalo, že student toho vyzkoušel více, ale z nějakého důvodu to nezahrnul do textu, pak bych byl ochoten hodnotit lépe, tedy „velmi dobře“ (B).

V Praze dne 16. 8. 2021

Ing. Jan Kamenický, Ph.D.