

Posudek oponenta bakalářské práce

Student: Šamánek Jan
Téma: Detekce výskytu objektů ve videozáznamu (id 23030)
Oponent: Orság Filip, Ing., Ph.D., UITS FIT VUT

- 1. Náročnost zadání** **obtížnější zadání**
Zadání je obtížnější především z hlediska nutnosti nastudování složitějších postupů analýzy videa, které nejsou běžnou součástí BSP.
- 2. Splnění požadavků zadání** **zadání téměř splněno**
Zadání v bodě 3 říká, že navržené řešení má umožnit detekci minimálně 5 různých objektů. Nicméně poskytnuté řešení klasifikuje scénu jako celek. Detekci objektů si představuji tak, že na vstupu poskytnu snímek a jako výstup obdržím polohu a typ objektu, což poskytnuté řešení zjevně nedělá a práce se nezmiňuje o tom, že by algoritmy hledaly polohu objektů, které by posléze klasifikovaly. Nevím, zda to byl záměr, či nikoliv, ale nepovažuji v tomto případě zadání za splněné a ponechám na případně obhajobě vysvětlení tohoto přístupu.
- 3. Rozsah technické zprávy** **je v obvyklém rozmezí**
- 4. Prezentační úroveň předložené práce** **65 b. (D)**
Logická struktura technické zprávy, stejně jako rozsah a návaznost kapitol jsou v pořádku. Nicméně pochopitelnost práce může snižovat snaha o široký záběr témat a chybí tak detailnější teoretický rozbor konkrétních použitých řešení. Z práce vyplývá, že autor vytvořil vlastní neuronovou síť, pro účely klasifikace snímků. Proč se tedy v práci zbytečně hovoří o metodách shlukování, random forests, Gaussovských modelech nebo o práci s barvami? Není zřejmé, jak tyto metody figurují v navrženém řešení.
- 5. Formální úprava technické zprávy** **50 b. (E)**
Z jazykového hlediska je práce velmi slabá. Autor zcela nechápe shodu podnětu s přísudkem, a i díky množství dalších gramatických chyb a nevhodné stavbě vět je celá práce těžko čitelná. Zápis matematických vzorců s centrováním slovem "kde" na samostatném řádku a výčtem, který následuje, spolu s přehnanou úrovní číslování také poněkud vybočuje ze zažitých pravidel.
- 6. Práce s literaturou** **85 b. (B)**
Student k vypracování práce použil dostatečné množství relevantních literárních zdrojů, které náležitě odcitoval. Naprostá většina zdrojů jsou online odkazy, dále několik článků a pouze minimum knih. Převzaté části jsou dostatečně odlišeny od vlastních úvah autora. Citace jsou většinou v souladu s citačními zvyklostmi.
- 7. Realizační výstup** **75 b. (C)**
Realizačním výstupem je sada skriptů v jazyce Python využívající různé knihovny (především pak Qt pro GUI, OpenCV pro práci s obrazem a Tensorflow pro práci s neuronovými sítěmi). Zdrojový kód je řádně komentovaný a (už z principu Pythonu) dobře formátovaný. Aplikační výstup splňuje zadání až na již zmíněnou nejasnost (detekce objektu vs. klasifikace celého snímku nebo vybraného výřezu). Ve zdrojových datech postrádám označovanou trénovací sadu, která je použita pro učení sítí.
- 8. Využitelnost výsledků**
Výsledné GUI a skripty lze použít pro jednoduché učení sítí klasifikujících snímky do několika tříd a pro testování jejich kvality. Pro širší využití by bylo nutné řešení rozšířit o další funkce.
- 9. Otázky k obhajobě**
 - Vysvětlíte, jaké objekty vaše řešení detekuje a v čem ta "detekce" spočívá. Z textu to není jednoznačně zřejmé a já si pod pojmem "detekce objektů" představuji něco jiného.
 - Jaké jsou vstupy pro trénování sítě (například formát, velikost, poloha objektu, značka třídy objektu)?
 - V čem konkrétně se liší vámi navržená síť od původní VGG-net?
 - Která síť v uvedených grafech je vámi vytvořená (například tabulka 5.1 uvádí pojem "vlastní model" v ostatních grafech nic takového není)?
 - Co znamená "přesnost" a "poměr správně označených snímků" v grafech z matematického hlediska?
- 10. Souhrnné hodnocení** **49 b. nevyhovující (F)**
Celkově bych práci ohodnotil sice jako slabší, ale dostatečnou. Celý text velmi oslabuje nízká jazyková kvalita a zvláštní formátování. Jinak je zpracování na průměrné úrovni a ničím zvláštním nevykíná. Nicméně nejsem si zcela jist, zda bylo správně pochopeno zadání ať už mnou nebo studentem. Rozcházejí se v tom, co je detekce objektů. Dle mého názoru by výstupem sítě měla být informace o tom, že byl nalezen objekt, jeho poloha a třída.

Z práce plyne, že výstupem je spíše klasifikace snímku jako celku, případně jeho výřezu - tedy nedetekují objekt, ale třídu snímku (v konkrétním případě uvedeném v práci jde například o sporty). Pokud je to jinak, pak to z práce není vůbec zřejmé a ponechám prostor na obhajobu a posouzení komise, zda je zadání splněno nebo ne. Pokud student uspokojivě vysvětlí celou situaci, lze práci ohodnotit stupněm D (60 bodů), jinak ji považuji za nedostatečnou a hodnotím ji stupněm F.

Prohlášení: Uděluji VUT v Brně souhlas ke zveřejnění tohoto posudku v listinné i elektronické formě.

V Brně dne: 24. června 2020

Orság Filip, Ing., Ph.D.
oponent