

# Posudek oponenta bakalářské práce

Autor práce: Filip MAJER

Název práce: Detekce objektů s využitím textových dotazů

## Jazyková a grafická úprava

Nadprůměrné

## Formální a obsahová stránka práce

Průměrné

## Vhodnost použitých metod

Nadprůměrné

## Způsob zpracování a vyhodnocení

Nadprůměrné

## Správnost získaných výsledků

Průměrné

## Vlastní přínos

Nadprůměrné

## Doplnění hodnocení, připomínky:

Předložená práce se věnuje problematice detekce vizuálních objektů pomocí textových dotazů. Jedná se o problematiku kombinace dvou modalit, související s porozuměním textu a obrazu. Práce je velice ambiciózní a absolutně v souladu s moderními technikami zpracování multimodálních dat v oboru umělé inteligence. Použité techniky jsou správně zvoleny a práce pouze nereplikuje známé experimenty, ale pohlíží na problém v novém světle. Nedostatky práce jsou pouze formální. Jako příklady uvedu: rovnice by měli být zakončeny tečkou, nebo čárkou, dle kontextu; Obrázek 1 uvádí jiné označení parametrů oproti textu; některé obrázky nejsou v textu referencovány; ve vzorcích 9 a 10 je chybně konstanta ve jmenovateli oproti citovaným zdrojům, atd. Jako největší formální nedostatek vnímám nedostatečné bibliografické údaje u citovaných prací, kdy často chybí název periodika a vydavatel. Tyto nedostatky lze jednoduše odstranit a nevnímám je jako kritické. Líbí se mi studentova diskuze o dosažených výsledcích i vyslovené hypotézy, ze kterých plyne uvedena budoucí práce. Přidal bych více ukázek výstupů systému, aby bylo vidět co funguje a naopak co selhává. Velice oceňuji publikovanou online aplikaci, ve které lze systém testovat.

## Dotazy

1. V práci uvádíte, že konvoluce je invariantní vůči translaci. To není správně. Konvoluce je ekvivariantní vůči translaci. Popište rozdíl mezi invariancí a ekvivariancí.
2. V analýze výsledků používáte pojem "skutečné výstupy". Vhodnější označení je "požadované výstupy". V angličtině Ground Truth. Jak by jste tento pojem vysvětlil?
3. F1 skóre je určitý typ průměru hodnot Precision a Recall. Jak se tento typ průměru jmenuje?
4. K diskuzi: Systém má problémy detekovat hlavně malé objekty. Souhlasím s Vaším návrhem použít menší patche na vstupu Vision Transformeru. Na druhou stranu, šlo by pomocí výřezání a zvětšení malých objektů rozšířit trénovací vzorky pro tyto objekty? Jaký vliv by to podle Vás mělo na detekci malých objektů v testovacích datech?

## Splnění bodů zadání

úplně

**Doporučení k obhajobě**

ANO

**Hodnocení: 1 - Výborně**

V \_\_\_\_\_ dne \_\_\_\_\_

-----  
Ing. Marek Hruz, Ph.D.