

doc. Ing. Petr Doležel, Ph.D.  
Katedra řízení procesů  
Fakulta elektrotechniky a informatiky  
Univerzita Pardubice

## **Oponentní posudek diplomové práce**

**Bc. Tomáš Vladyka**

### **Robotické vidění s využitím umělé inteligence**

Cílem předložené diplomové práce je vytvoření vlastního systému robotického vidění s prvky umělé inteligence, který bude schopen lokalizovat a kategorizovat předem naučené objekty. Praktickým výstupem je zejména soubor vyvinutých softwarových nástrojů pro návrh proprietárního systému robotického vidění, ale také podrobná dokumentace popisující činnosti vykonané autorem práce.

Práce je kromě úvodu a závěru rozdělena do dvou komplexních kapitol, přičemž první z nich se věnuje teoretickým podkladům pro tvorbu práce a druhá se zaměřuje na věcný popis aktivit vykonaných autorem.

Rešeršní činnosti, kterým se autor věnuje na dvaceti stranách, poskytují úvod do problematiky strojového vidění a konvolučních neuronových sítí, v poslední části úseku jsou představeny systémy 2D a 3D robotického vidění. Tato část práce je psána stručně, přehledně a srozumitelně.

Z hlediska praktické stránky práce autor sice poskytl přehledný a uspořádaný soubor informací o realizovaných krocích, nicméně pro korektní uchopení práce chybí komplexní stanovení cílů. Autor pouze vágně definoval celkový cíl jako „vytvoření vlastního systému robotického vidění s prvky umělé inteligence, který bude schopen lokalizovat a kategorizovat předem naučené objekty“ a dále se věnoval sekvenčnímu popisu jednotlivých provedených kroků. Na základě textu není možné rozhodnout o míře naplnění cílů ani o náročnosti řešení, pokud definována explicitní kritéria naplnění typu jaké objekty mají být lokalizovány a kategorizovány, jaká je požadovaná přesnost, či požadavky na výpočetní čas. Autor některé indikátory diskutuje v místě popisu návrhu řešení bez jakékoliv vazby na dosažení cílů, celá práce tak částečně působí jako cvičení spíše než jako uzavřené dílo.

Řešení autor práce rozdělil do několika na sebe navazujících sekcí. Nejprve definoval soubor elementárních funkcí pro implementaci nástrojů hlubokého učení. Zde je důležité zdůraznit, že autor nevyužil žádný z dostupných frameworků a nástroje vyvinul sám. Pedagogický dopad tohoto rozhodnutí je vysoký, na druhou stranu tím autor ztratil možnost vytvářet složité výpočetně optimalizované struktury využitelné v dalších krocích. Následně se věnoval problému lokalizace objektu v obrazových datech, kdy se rozhodl využít klasických nástrojů zpracování obrazu. Díky tomuto rozhodnutí však vyvstalo omezení na definované pozadí v rámci snímku. Pokud by např. byla využita architektura R-CNN popisovaná v teoretické části práce, toto omezení by bylo překonáno. Následně autor popisoval řešení klasifikace lokalizovaných objektů pomocí několika jednoduše navržených modelů založených na architekturách konvolučních neuronových sítí. Bohužel v tomto kroku nedocházelo k žádné systematické optimalizaci parametrů ani hyperparametrů použitých modelů.

Bez ohledu výše zmíněné nedostatky lze konstatovat, že předložené řešení je korektně prezentováno a přesnost výsledných modelů je dobře otestována.

Po formální stránce je práce na výborné úrovni. Je psána srozumitelně a přehledně, nevšiml jsem si žádných stylistických chyb. Z pohledu typografie lze vytknout prezentaci kódu formou bitmapových obrázků a kvalita grafických výstupů by obecně mohla být vyšší.

Pro obhajobu práce mám následující dotazy a požadavky.

- Vysvětlíte, proč pro lokalizaci objektů nebyla použita architektura R-CNN, případně jiná moderní architektura (YOLO, SSD). Předěšlo by se tím omezení na definované pozadí v jednotlivých snímcích.
- Obecně pro testování byly použity datasety o velmi malém rozsahu. Byl ještě nějaký jiný důvod, proč autor zvolit takto malé datasety, kromě časové náročnosti spojené s přípravou? V rámci práce by bylo vhodné testovat např. vliv rozsahu datasetu na výslednou přesnost modelu.
- Při testovacím scénáři zaměřeném na klasifikaci dvou podobných objektů nebylo dosaženo nijak pozoruhodné přesnosti. Proč autor více neexperimentoval s použitou architekturou modelu? V praxi se např. používají architektury s významně vyššími počty použitých filtrů v konvolučních vrstvách, čímž by se mohlo dosáhnout přesnosti použitelné v praxi.
- Pro práci jsou relevantní pojmy *počítačové vidění (computer vision)*, *strojové vidění (machine vision)* a *robotické vidění (robotic vision)*. Je mezi nimi nějaký rozdíl? Diskutujte v rámci obhajoby.

Závěrem mohu konstatovat, že předložená práce splňuje požadavky na práci tohoto typu a student splnil cíle, jak byly definovány v zadání práce. Vzhledem k faktům uvedeným výše při úspěšné obhajobě a zodpovězení všech dotazů navrhuji známku

=C=

26. 5. 2024

Petr Doležel