

Posudek bakalářské práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce	Jakub Hajko	
Název práce	Comparison of joint-embedding deep networks for known-item search in image datasets	
Rok odevzdání	2024	
Studijní program	Informatika	
Specializace	Umělá Inteligence	
Autor posudku	Ladislav Peška	Oponent
Pracoviště	Katedra softwarového inženýrství	

K celé práci

lepší OK horší nevyhovuje

	lepší	OK	horší	nevyhovuje
Obtížnost zadání		X	X	
Splnění zadání		X		
Rozsah práce <i>... textová i implementační část, zohlednění náročnosti</i>		X		
<p>Cílem práce bylo srovnat state-of-the-art modely generující joint text-image embeddingy (zejména varianty CLIP modelu) na retrieval úlohách. Práce se zabývá dvěma variantami tasku: text-based retrieval pro known-item search a konzistence vypočtené image-to-image similarity s podobností vnímanou lidmi. Práce je především experimentálního charakteru. Autor vybral několik předtrénovaných modelů a otestoval je na 3 datasetech pro text-based retrieval task a jednom dalším pro konzistenci vnímání podobnosti. Pro vybrané snímky z text retrieval datasetů autor připravil dvě varianty queries (krátké a delší). V případě tasku na vnímání podobnosti pak autor postupoval na základě evaluačního protokolu z related work, který rozšířil o vybrané state-of-the-art modely.</p> <p>Experimentální výsledky jsou zajímavé a přispívají k současnému stavu poznání ve zvolené oblasti. Výsledky lze poměrně přímočaře aplikovat ve vyhledávacích soutěžích, jako je VBS nebo LSC.</p>				

Textová část práce

lepší OK horší nevyhovuje

	lepší	OK	horší	nevyhovuje
Formální úprava <i>... jazyková úroveň, typografická úroveň, citace</i>		X		
Struktura textu <i>... kontext, cíle, analýza, návrh, vyhodnocení, úroveň detailu</i>		X		
Analýza		X	X	
Vývojová dokumentace		X		
Uživatelská dokumentace		X		

Text práce je po jazykové stránce solidní a dobře strukturovaný. V některých částech bych si však dovedl představit vícero detailů. Například v introduction autor opomenul vysvětlit, co je vlastně cílem práce. Také kapitola 2 by místy snesla více podrobností - například jen poměrně povrchně je vysvětlen rozdíl mezi ALIGN a CLIP architekturami.

Trochu slabší je u provedení ne zcela dostatečné odůvodnění zvolených experimentálních variant, které zabraňuje pořádně provést ablační studii. Např., autor použil varianty CLIP modelu předtrénovaného nad 3 různými datasey, ale tyto modely také zahrnovaly jiné encoder varianty, takže i když autor v závěru tvrdí, že modely trénované na webli datasetu dosahovaly nejlepších výsledků, z experimentů není zcela jasné, zda to je efekt samotného datasetu, nebo odlišností ve zvolených architekturách (např. jiné verzi vision transformeru).

Implementační část práce

lepší OK horší nevyhovuje

Kvalita návrhu ... <i>architektura, struktury a algoritmy, použité technologie</i>		X		
Kvalita zpracování ... <i>jmenné konvence, formátování, komentáře, testování</i>		X		
Stabilita implementace		X		

Samotná implementace - jak nástroje pro datovou augmentaci, tak i experimentální evaluace a analýza výsledků jsou odpovídající a umožňují vcelku snadné budoucí rozšíření experimentů.

Celkové hodnocení Výborně (spíše horší)
Práci navrhuji na zvláštní ocenění Ne

Datum

Podpis