

# text-to-image search/ image-to-image search\*

Ondrej Krajčovič

Slovenská technická univerzita v Bratislave

Fakulta informatiky a informačných technológií

xkrajcovico@stuba.sk

5.november 2023

## Abstrakt

Článok poskytuje stručný opis fungovania, využitia a možností spätného vyhľadávania obrázkov (eng.: reverse image search, alebo image-to-image-search) a vyhľadávania obrázkov pomocou textu (eng.: text-to-image search). Nakoľko ľudia prirodzene vyjadrujú svoje myšlienky slovami, je pre nás pomerne ľahké komunikovať pomocou obrázkov, obzvlášť ich vyhľadávanie je občas pomerne neefektívne. Komunikácia sa však v posledných rokoch veľmi zmenila nástupom moderných, technologicky vyspelých komunikačných kanálov(eng.: Mass media). Vyjadrovanie myšlienok graficky sa teda zdá byť na dosah ruky, problémom je však vyhľadanie daných grafík v databázach a na internete. Článok rozoberá nedávny pokrok, ľahkosti a nové nápady, ktoré spájajú rôzne typy informácií, predovšetkým vizuálne a modálne, a ukazuje, aké dôležité sú tieto oblasti v dnešnej digitálnej dobe.

---

\*Semestrálny projekt v predmete Metódy inžinierskej práce, ak. rok 2023/24, vedenie: Vladimír Mlynarovič

## 1 Úvod - Ľudské vnemy a sémantika

Ľudský mozog vníma informácie pomocou zmyslov. Dominantný zmysel u človeka je zrak, je nám preto iba prirodzené vnímať informácie pomocou obrázkov. Ľudská schopnosť sémanticky pochopiť obraz je dokonca oveľa rýchlejšia ako schopnosť vnímať text. Toto je presne dôvod, prečo sú dôležité informácie často znázorňované pomocou rôznych graffík, piktogramov 1. Rôzne abecedy sa dokonca vyvýiali priamo z grafických znázornení objektov ktoré sa vyskytovali v okolí osôb vyvýjajúcich dané písmo. Pre počítače však vizuálny spôsob vnímania komplexných informácií nie je prirodzený. Ako teda vnímajú stroje informáciu?



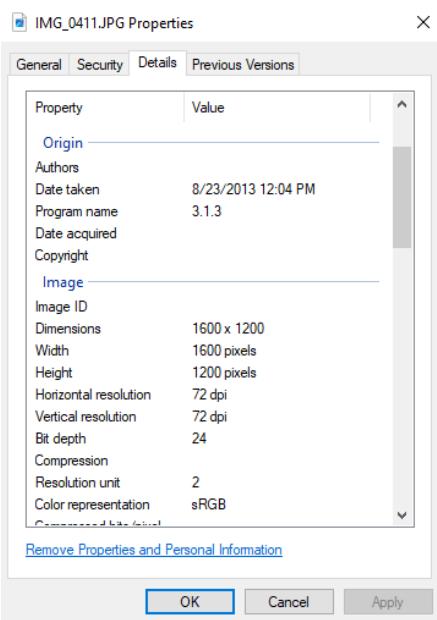
Obr. 1: ukážka piktogramu

## 2 funkcia metadát (metaúdajov) pri hľadaní obrázkov

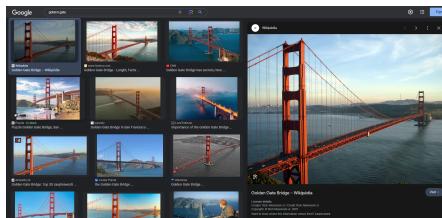
Metadáta sú všeobecne informácie, respektíve dáta, ktoré nemusia byť priamo súčasťou daného súboru, ale ich primárной funkciami je opísat, čo sa v danom súbore nachádza. V našom prípade sú to teda informácie slovne opisujúce čo sa na obrázku nachádza, kde a kedy bol obrázok vytvorený (tieto informácie platia predovšetkým o fotografiách) 2. Tento kontext teda odstraňuje rozdiel medzi jazykovým a vizuálnym obsahom a umožňuje získať presnejšie a relevantnejšie výsledky vyhľadávania. Slúžia teda ako kľúč ku spracovaniu vizuálnych informácií, napríklad pomocou queries, čo znamená, že vyhľadávací algoritmus vyhľadáva na základe nie iba grafických, ale predovšetkým textových informácií. Takéto vyhľadávanie sa nazýva text to image search.

## 3 Text-to-image search

Queries, textový vstup poskytnutým používateľmi sa pri tomto procese porovnáva z potencionálne zhodnými informáciami zapísanými vo forme metadát k prislúchajúcim obrázkom. V prípade že do vyhľadávača na internete zadáme vstup: "golden bridge", pomocou metadát vyhľadá súvisiace obrázky, ktorých metadáta obsahujú text: " golden bridge" 3



Obr. 2: ukážka zobrazenia metadát na konkrétnom obrázku v prostredí windows



Obr. 3: ukážka využitia metadát pri hľadaní na internete

#### 4 Neefektívnosť pri vyhľadávaní obrázkov

V prípade, že však pri vyhľadávaní chceme hľadať pomocou vizuálneho vstupu, a nie pomocou textových queries, musíme postupovať inak. Jednou z možností je preložiť grafickú informáciu, ktorú nám poskytuje vstupný obrázok do textovej podoby, a následne pomocou text-to-image vyhľadáva. Tento proces je však pomerne neefektívny, a jeho výsledky mnohokrát nezodpovedajú našim očakávaniam natoľko aby sme s nimi boli spokojní. Optimálnym riešením tohto problému by teda bol počítač, schopný vnímať a pracovať s informáciami priamo na úrovni grafického rozlíšenia týchto informácií.

## 5 image-to-image search

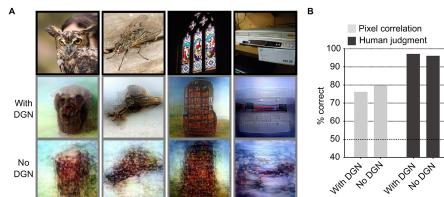
Je to vyhľadávanie obrázkov kedy je vstupom(query) iný obrázok. Umožňuje používateľom nájsť obrázky podobné danému vyhľadávaciemu obrázku, často analyzovaním vizuálnych prvkov, ako sú farba, tvar a textúra

## 6 Reverse Image Search

Algoritmus používaný na image-to-image search. pracuje v nasledovných krokoch: najprv vyhľadáva špecifické črty obrázka, ďalej porovnáva jednotlivé črty s črtami iných obrázkov v databáze (napríklad na internete). Nakoniec vyhodnotí najpodobnejšie obrázky a vráti ich používateľovi. počas tohto procesu nemusí algoritmus používať iba črty daného obrázka, môže si pomáhať aj metadátami

## 7 Sémantika

Vedci boli schopní graficky zobraziť [?] neurónové signály v ľudskom mozgu priamo pri procese spracovávania vizuálnych podnetov 7, a následne aj počas toho ako si daný človek spomína na daný vizuálny podnet. Výsledkom nebol jasný obraz, zhodný s vizuálnym podnetom, ktorý bol človeku poskytnutý, ale pomerne rozmazená a skreslená informácia. Mňa osobne zaujala podobnosť tejto grafiky s obrázkami ktoré generovala umelá inteligencia.



Obr. 4: Vizuálne zobrazenie signálov z ľudského mozgu pri spracovávaní obrázkov

## 8 AI a deep learning

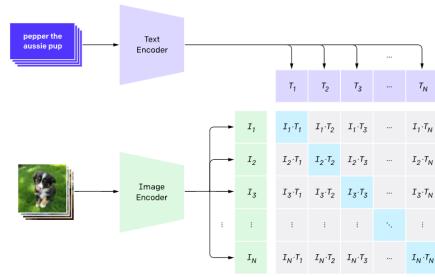
Spoločnosť OpenAI publikovala v roku 2021 nástroj s názvom CLIP(Contrastive Language-Image Pre-Training 8), Je to pomerne efektívna umelá inteligencia 8 nástroj, ktorý môže generovať textové opisy obrázkov, hľadať obrázky založené na textových queries a vykonávať rôzne úlohy, ktoré si vyžadujú hlboké pochopenie obrazov aj textu

## 9 Záver

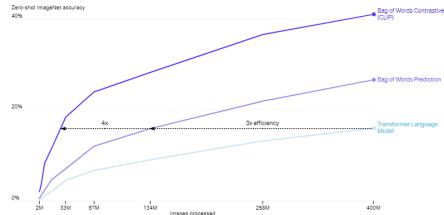
Článok sa dotýka tém vyhľadávania, s hlavným zameraním "text-to-image search a image-to-image search". Je v ňom vysvetlená dôležitosť a funkcialita



Obr. 5: Obrázok generovaný pomocou umelej inteligencie



Obr. 6: CLIP spracováva vizuálne informácie a metadáta pri trénovaní



Obr. 7: efektivita nástroja CLIP

metadát pri vyhľadávaní. Koncept image-to-image search vyhľadávania je čitateľovi taktiež priblížený. Na záver sa článok dotýka témy umelej inteligencie a jej konkrétneho využitia v tejto oblasti

## **Literatúra**