

SISTEM ZA REVERZNU PRETRAGU SLIKA**REVERSE IMAGE SEARCH SYSTEM**Ivan Mršulja, *Fakultet tehničkih nauka, Novi Sad***Oblast – PRIMENJENE RAČUNARSKE NAUKE I INFORMATIKA**

Kratak sadržaj – Ovaj rad predlaže alternativni pristup reverznoj pretrazi slika korišćenjem modela mašinskog učenja za izvlačenje tagova i svojstava iz slike. Svođenjem vektorske pretrage na pretragu teksta, predloženi metod povećava efikasnost i tačnost pretrage slika.

Ključne reči: Reverse pretraga slika, mašinsko učenje, računarska vizija.

Abstract – This paper proposes an alternative approach to reverse image search by leveraging machine learning models to extract tags and features from an image. By reducing vector image search to a full-text search, the proposed method improves the efficiency and accuracy of image retrieval.

Keywords: Reverse image search, machine learning, computer vision.

1. UVOD

Reverzna pretraga slika (*reverse image search*) je tehnika pretrage koja omogućava korisnicima da pronađu slične ili identične slike na internetu koristeći već postojeću sliku kao ulazni parametar [1]. Umesto da korisnici unose tekstualne upite, mogu otpremiti sliku ili uneti URL slike kako bi pronašli druge slike sa sličnim vizuelnim karakteristikama [2]. Algoritmi za reverznu pretragu slika analiziraju ključne elemente slike, kao što su boje, oblici, teksture ili drugi vizuelni detalji, kako bi pronašli slike sa sličnim karakteristikama iz baza podataka ili veb stranica širom interneta. Ova tehnika je korisna u mnogim situacijama, kao što su pronalaženje izvora slike, identifikacija objekata, pronalaženje sličnih vizuelnih sadržaja, otkrivanje plagijata ili pronalaženje više informacija o određenom objektu na slici. Reverzna pretraga slika može biti od pomoći u istraživačkim, kreativnim i bezbednosnim kontekstima, pružajući korisnicima mogućnost da pronađu relevantne slike na osnovu već postojeće vizuelne reference [2].

U ovom radu, rešavan je problem implementacije jednog ovakvog sistema, gdje korisnik može unijeti proizvoljnu sliku kao uzorak upit, a sistem mu za isti vraća slike koje najbolje odgovaraju na njega. Pored toga, omogućeno je i indeksiranje slika. Takođe, adresiran je i problem velikog broja korisnika koji bi potencijalno koristili ovaj

sistem, te se razvoj vodio praksama servisno-orijentisane arhitekture gdje je svaki servis moguće nezavisno skalirati u proizvoljnom obimu.

1.1. Svođenje problema pretrage vektora na pretragu teksta

Jedinstvenost ovog rješenja leži u formatu svojstava koja se koriste prilikom indeksiranja i pretrage. Većina današnjih sistema se oslanja na vektorske reprezentacije dokumenata kako bi implementirala pretragu. Iako je ovo veoma efikasan način pretrage vizuelno sličnih slika, veliki je problem što se eksplicitno u obzir ne uzimaju svi objekti sa slike već se čitava slika reprezentuje kao neki *black-box* skup svojstava.

Sa druge strane, moje rješenje svodi pretragu slika na pretragu teksta, modelom mašinskog učenja (*YOLOv5*) se prepoznaju objekti, koji se posmatraju kao ključne riječi prilikom indeksiranja u *Elasticsearch* sistemu za pronalaženje informacija, pomoću kojih možemo prepoznati sve kontekstualno slične slike sa uzorkom dok se prostor boja slika kao opcioni filter kako bismo mogli dobiti vizuelno najbližnje slike.

1.2. Elasticsearch

Elasticsearch je visoko skalirajući, distribuirani sistem za pretragu i analizu podataka. On je osmišljen da brzo i efikasno indeksira, čuva i pretražuje velike količine strukturisanih i nestrukturisanih podataka. *Elasticsearch* je baziran na *Apache Lucene* biblioteci, koja pruža moćne algoritme za pretragu. Takođe podržava brojne napredne funkcionalnosti kao što su pretraga punim tekstom (*full-text search*), filtriranje, agregacija, geolokacija, analiza teksta, vektorsko indeksiranje i pretraga i još mnogo toga.

Osim pretrage, *Elasticsearch* takođe omogućava analizu podataka i vizualizaciju rezultata. Kroz *Kibana* interfejs, korisnici mogu pristupiti bogatim vizualizacijama, izveštajima i nadzoru nad podacima. *Elasticsearch* je dizajniran da bude otporan na kvarove i pruža mogućnost replikacije podataka i distribucije preko više čvorova.

Takođe podržava horizontalno i vertikalno skaliranje, omogućavajući dodavanje novih čvorova kako bi se postigla bolja brzina i otpornost na visoka opterećenja, kao i unapređenu dostupnost usluge. Sa svojom fleksibilnošću, skaliranjem i moćnim alatima za pretragu i analizu, *Elasticsearch* je postao popularan izbor za razne primene kao što su pretraga veb stranica, analiza logova i sistemi za praćenje događaja (*event tracking systems*) [4].

NAPOMENA:

Ovaj rad proistekao je iz master rada čiji mentor je bio dr Dragan Ivanović, red. prof.

1.3. YOLOv5

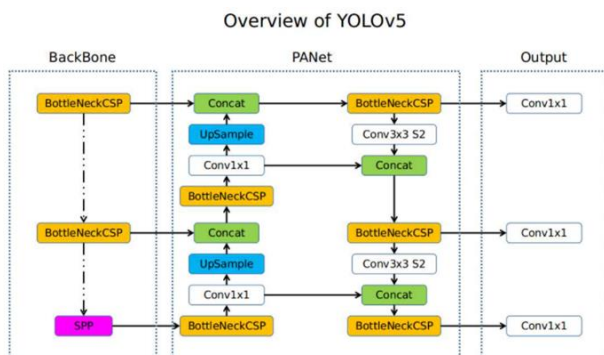
YOLOv5 model za detekciju objekata baziran je na *end-to-end (e2e)* arhitekturi i sastoji se od tri glavne komponente:

1. Vlasničke mreže (*backbone network*)
2. Glavne mreže (*neck network*)
3. Mreže izlaza (*output network*)

Vlasnička mreža (*backbone*) u YOLOv5 modelu je CSPDarknet53 mreža, koja se bazira na konvolucionim slojevima. Ova mreža je modifikovana verzija originalne Darknet53 mreže i pruža bolju reprezentaciju svojstava i razumevanje konteksta objekata. Ona je odgovorna za identifikaciju i izdvajanje bitnih svojstava iz ulaznih slika.

Glavna mreža (*neck network*) je dodatni sloj između vlasničke mreže i mreže izlaza. Ona ima za cilj da usavrši svojstva iz vlasničke mreže i podese ih tako da bi lakše detektovala objekte različitih veličina i proporcija. U specifikaciji YOLOv5 modela, koristi se PANet mreža koja je korišćena i u starijem YOLOv4 modelu.

Mreža izlaza (*output network*) predstavlja poslednju komponentu modela i sadrži konvolucione slojeve koji izvode konačne predikcije o klasama i lokacijama objekata. Ona preobrađuje izlazne "feature" iz glavne mreže u *anchor-box*-ove i vraća koordinate, sigurnost i klase detektovanih objekata za svaki *bounding-box*. U pojedinoj literaturi, ovaj segment u "pipeline"-u modela naziva se i "head network" [5]. Na slici 1 prikazana je gore opisana arhitektura.



Slika 1. Arhitektura YOLOv5 modela

1.4. Pregled relevantne literature

U radu [3] prikazana je *Similar Looks* funkcionalnost unutar *Pinterest* aplikacije koja omogućava korisnicima da izaberu predmet unutar slike, stavljajući okvir oko njega, a zatim vraća vizuelno slične slike/predmete. Glavni cilj ove funkcije je pomoći korisnicima da pronađu stvari koje ne mogu imenovati [2].

Indeksiranje slika se u ovom sistemu vrši po *incremental fingerprinting service (IFS)* sistemu. Ovaj sistem funkcioniše po principu da svaka slika ima svoj otisak koji se sastoji od svih odvojenih objekata koji se nalaze unutar te slike (npr. cipele, torbe, čaše itd.). Svaki objekat ima oznaku koja specifikuje ime klase kojoj objekat pripada [3].

Prilikom pretrage, sistem koristi tehnologiju prepoznavanja objekata da lokalizuje i klasifikuje predmete (objekte)

na slikama. Iz ovako detektovanih objekata se izvlače karakteristike i koriste se za dalju pretragu. Drugim riječima, skraćuje se zadatak detekcije objekata više klasa u klasifikaciju kategorija, tako da umesto da se traži podudaranje u svim slikama na *Pinterest-u*, prvo se dobavljaju slike koje su u istoj kategoriji.

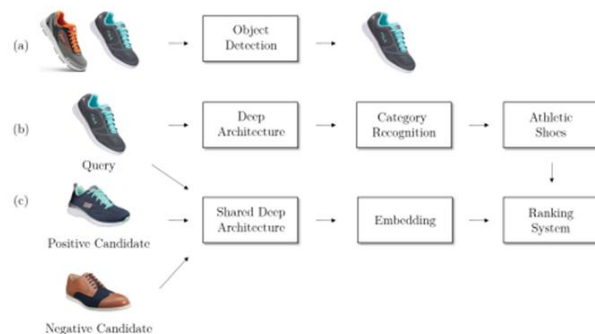
Ovo je odličan pristup jer se upitna slika upoređuje sa slikama koje imaju visoku vjerovatnoću da su slične. Prethodni korak filtriranja povećava stopu tačnih pozitiva. Sistem se zasniva na konceptima dubokog učenja, kao što su konvolutivne neuronske mreže (*CNN*) i detektori objekata zasnovani na dubokom učenju (*Single Shot Detectors, SSD*).

Sistem je implementiran korišćenjem radnog okvira *Caffe* (*Java* tehnologija). Podaci sistema se skladište na *Amazon S3-u* [2][3].

U radu [1] tok rada vizualne pretrage je sledeći: korisnik *upload*-uje sliku ili je snima kamerom, a rezultat će biti slične slike i predmeti; korisnik može izabrati da li želi da kupi (slično "*Shopping*" sekciji kod *Google* pretraživača) ili dalje istražuje predmete [2]. Ovaj sistem obuhvata tri glavne faze:

1. Obuka modela: U sistemu se koristi nekoliko modela dubokih neuronskih mreža kako bi se poboljšala relevantnost rezultata, kao što su *AlexNet*, *ZFSPPNet*, *GoogleNet*, *ResNet*, a takođe se koristi i zajednički algoritam *kmeans* kako bi se izgradio inverzni indeks u nivou-0 podudaranja, pri čemu nivo 0 označava najmanju sličnost sa upitnom slikom [1].
2. Razumevanje upita (slike): Iz upitne slike izvlače se različite karakteristike koje opisuju njen sadržaj, uključujući karakteristike prepoznavanja kategorije, prepoznavanja lica, boje i otkrivanja duplikata, enkodere dubokih neuronskih mreža (*DNN*) i detekciju objekata (slika 2).
3. Pretraga slika: Pronalazi vizualno slične slike na osnovu izvučenih karakteristika i namere korisnika.

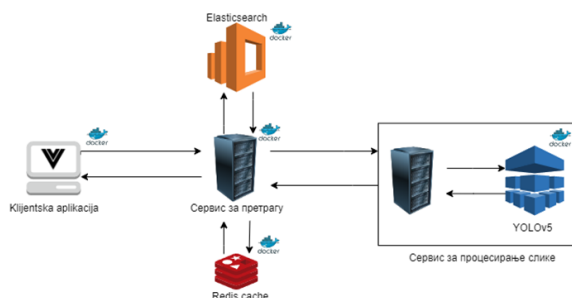
Svi modeli se treniraju na obučavajućem skupu podataka koji su prikupljeni za određene domene npr. kupovina. Za nadzor obuke *DNN* koriste se više funkcija gubitka, kao što su *Softmax*, *Pairwise-loss* i *Triplet-loss* funkcije [1].



Slika 2. Primjene DNN modela koji se koriste u Bing-ovom sistemu

2. SPECIFIKACIJA SISTEMA ZA REVERZNU PRETRAGU SLIKA

U ovom poglavlju prikazana je arhitektura sistema sa aspekta dijagrama razmeštaja. Dijagram razmeštaja (slika 3) prikazuje softverske komponente i veze uz pomoć kojih iste komuniciraju.



Slika 3. Dijagram razmeštaja

Ovo softversko rješenje se sastoji iz dva zasebna servisa:

1. Servisa za pretragu i indeksiranje implementiranog u *Java Spring Boot* radnom okviru
2. Servisa za procesiranje slike implementiranog u *Python FastAPI* radnom okviru uz oslonac na *OpenCV* i *NumPy* biblioteke za implementaciju i integraciju neophodnih algoritama mašinskog učenja koji se koriste prilikom procesiranja slike

Pored ova dva servisa bitno je napomenuti i da postoji odvojeni frontend sloj implementiran u *Vue3* radnom okviru koristeći *JavaScript* programski jezik.

Servis za pretragu i indeksiranje komunicira sa servisom za procesiranje slike kako bi izvukao svojstva (*features*) sa slike koja je data kao uzorak za upit.

Pored funkcije pronalaženja informacija i indeksiranja, ovaj servis vrši i keširanje svojstava slike koristeći *Redis key-value* bazu podataka kako bi ubrzao vrijeme pretrage za više sukcesivnih upita sa istom slikom (pri listanju paginiranih rezultata), što predstavlja veoma bitan faktor u radu ovog sistema kako bi bio pogodan za korišćenje. Indeksiranje kao i pretraga je implementirana koristeći *Elasticsearch* sistem za pronalaženje informacija.

3. EKSPERIMENT

Eksperiment je sproveden u vidu ankete:

1. Svaki učesnik prvo instalira aplikaciju koristeći *docker-compose* alat.
2. Nakon toga, svaki korisnik treba da indeksira indeksni skup podataka kroz formu za indeksiranje.
3. Na kraju, za svaku od slika u upitnom skupu podataka, korisnik ima polje u anketi da ocijeni ocjenom od 1 do 5 odziv, redoslijed odgovora, kao i redoslijed odgovora prilikom korišćenja opcije za sortiranje pomoću prostora boja.

Prilikom davanja ocjena, korisnik ima na uvid cijeli indeksni skup podataka kako bi mogao sam da donese odluku o broju *true positive* instanci u datom skupu podataka.

1.3. Skup podataka

Testiranje se vrši upotrebom ručno sakupljenog skupa podataka korišćenjem *Google Images* servisa. Sakupljeno je 30 slika za indeksni skup i 10 slika za upitni skup. Bitno je napomenuti da su slike izabrane tako da postoji veliki broj slika koje mogu da odgovore na više od jednog upita kako bi konačna procjena svih metrika bila ostavljena subjektivnom osjećaju korisnika.

1.4. Evaluacija

Evaluacija sistema se vrši na veoma jednostavan način. Korisnicima je naznačeno u upitniku da po sopstvenom nađenju izvuku mjeru odziva i redoslijed odgovora za svaki od *sample* upita.

4. REZULTATI I TUMAČENJA

Rezultati prethodno opisanog eksperimenta predstavljeni su u tabeli 1. Za svaku uzoračku sliku izvučene su sledeće statistike: prosječan odziv kod svih ispitanika, prosječna preciznost bez korišćenja dodatnih filtera, prosječna preciznost pri korišćenju sortiranja u prostoru boja. U ispitivanju je učestvovalo 10 kandidata.

Tabela 1. Rezultati eksperimenta

Sadržaj slike	Odziv	Preciznost (default)	Preciznost (HSV sort)
auto	5	3	3,25
torta	4	3,25	3
mačka	4,5	4,25	4,25
pas	3	4,25	4
igra u pijesku	4,75	2,75	4,25
pica	2,75	2	4,75
ptice	5	5	4,25
televizor	3,25	4	4
ulica	4,5	5	4
pribor za jelo	3,75	4,25	4,25
Prosjek	4,05	3,775	4

Iz priloženih rezultata može se zaključiti da je odziv sistema zadovoljavajuć sa prosječnom ocjenom 4,05 od 5, dok je redoslijed odgovora ocijenjen sa 3,775 i 4, zavisno da li je rezultat sortiran pomoću ugrađenog value parametra ili uz pomoć prostora boja. Najlošiji odziv je dobijen prilikom upita u kojem je kao uzorak korišćena slika sa picom. Kada je ovaj slučaj detaljnije analiziran, došlo se do nalaza da je pri procesiranju slike, pica detektovana kao kolač, stoga nije vraćena velika količina

slika na kojima se nalazio relevantan objekat. Način da se ovaj problem riješi je dodatno dotreniranje *YOLOv5* modela nad klasom objekta koja ima najviše problema prilikom klasifikacije.

5. ZAKLJUČAK

U ovom radu vidjeli smo jedan način implementacije sistema za reverzno pretraživanje slika. Ovakav sistem omogućava korisnicima da umesto da se oslanjaju na ključne reči ili opise za pretragu slika, na efikasan način pronađu željene informacije na osnovu same slike. Ovo može da ima raznoliku primenu, od pomoći u identifikaciji različitih predmeta i lokacija, do potpunog unapređenja iskustva kupovine preko interneta.

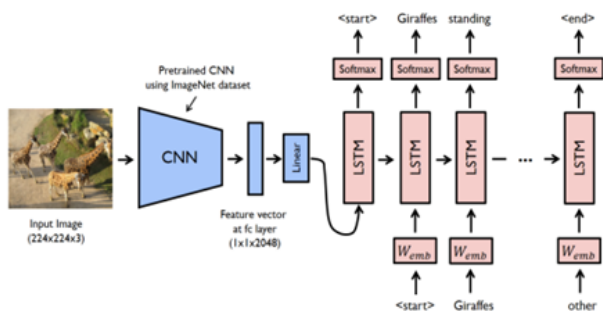
Sistem se sastoji iz dva odvojena servisa: servisa za pretragu koji vrši indeksiranje i pretragu koristeći *Elasticsearch* sistem za pronalaženje informacija, kao i keširanje koristeći *Redis in-memory* bazu, i servisa za procesiranje slike koji koristi moćne modele mašinskog učenja kako bi izvukao svojstva iz slike i predstavio je u formatu koji je lako pretraživ.

Za razliku od većine sistema koji vrše pretragu koristeći vektorski način pretrage (npr. *Web-Scale Responsive Visual Search*), ovaj sistem svodi problem pretrage slike na problem pretrage teksta gde se svojstva izvlače automatski, nasuprot tradicionalnom pristupu gde su se unosila ručno. Ovaj inovativan pristup već dugo je zastupljen u *Pinterest*-ovom *Similar Search* servisu i veoma lako može biti nadopunjen ručnim unosom opisa slike, ključnih reči itd.

5.1. Dalji razvoj sistema

S obzirom na to da je implementiran samo podskup različitih mogućnosti procesiranja slika, dodavanjem bilo kakvih dodatnih načina procesiranja moguće je postići poboljšanje sistema. Jedna stvar koja najviše ima smisla jeste da se postojeći način pretrage ukombinuje sa vektorskim načinom pretrage, tako da postojeći način isfiltrira kontekstno relevantne rezultate dok vektorska pretraga pronalazi vizuelno najsličniju sliku. Olakšavajuća okolnost kod ovog pristupa je ta što *Elasticsearch* podržava vektorsko indeksiranje i pretragu *out-of-the-box* [6].

Takođe, dodavanje modula za automatsko generisanje opisa ili ključnih reči iz slike korišćenjem mašinskog učenja moguće je poboljšati kvalitet izvučenih svojstava i proširiti domen primene gde bi se pored reverzne pretrage mogla implementirati i standardna tekst bazirana pretraga.



Slika 4. Prijedlog arhitekture pipeline-a za generisanje opisa slike

Jedan od načina na koji bi ovo moglo biti uvedeno jeste korišćenjem *CNN*-a za ekstrakciju obeležja (slika 4) i rekurentnu neuronsku mrežu (poput *LSTM*-a) ili transformer za generisanje teksta [7].

Još jedan zanimljiv način implementacije jeste integracija ovog sistema sa nekim *LLM*-om koji može generisati opis/ključne reči slike (na primer, *ChatGPT4*).

6. LITERATURA

- [1] Hu, H., Wang, Y., Yang, L., Komlev, P., Huang, L., Chen, X., Huang, J., Wu, Y., Merchant, M. and Sacheti, A., 2018, July. Web-scale responsive visual search at bing. In Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining (pp. 359-367).
- [2] Al-Lohibi, H., Alkhamisi, T., Assagran, M., Aljohani, A. and Aljahdali, A.O., 2020. Awjedni: a reverse-image-search application. *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal*, 9(3), p.49.
- [3] Jing, Y., Liu, D., Kislyuk, D., Zhai, A., Xu, J., Donahue, J. and Tavel, S., 2015, August. Visual search at pinterest. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1889-1898).
- [4] Elasticsearch, B.V., 2018. Elasticsearch. software, version, 6(1).
- [5] Zhu, X., Lyu, S., Wang, X. and Zhao, Q., 2021. TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 2778-2788).
- [6] Amato, G., Bolettieri, P., Carrara, F., Falchi, F. and Gennaro, C., 2018, June. Large-scale image retrieval with elasticsearch. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (pp. 925-928).
- [7] Yadav, A., Vishwakarma, A., Panickar, S. and Kuchiwale, S., 2020. Real time video to text summarization using neural network. *Int. Res. J. Eng. Tech*, 7, pp.1828-36.

Kratka biografija:



Ivan Mršulja rođen je 31.01.2000. godine u Kotoru. Smjer softversko inženjerstvo i informacione tehnologije na Fakultetu tehničkih nauka u Novom Sadu upisao je 2018 godine. Osnovne studije završio je u septembru 2022. godine. Od oktobra 2022. godine upisuje master studije i angažovan je kao saradnik u nastavi.

Kontakt: ivan.mrsulja@uns.ac.rs