

# Usporedba gotovih modela za pronalaženje semantički sličnih medicinskih slika / Comparison of of-the-shelf models for medical image retrieval

---

Tus, Ivana

Undergraduate thesis / Završni rad

2022

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Rijeka, Faculty of Engineering / Sveučilište u Rijeci, Tehnički fakultet**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:190:856364>

Rights / Prava: [In copyright](#) / [Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2024-09-02**



Repository / Repozitorij:

[Repository of the University of Rijeka, Faculty of Engineering](#)



SVEUČILIŠTE U RIJECI

TEHNIČKI FAKULTET

Preddiplomski sveučilišni studij računarstva

Završni rad

**USPOREDBA GOTOVIH MODELA ZA PRONALAZENJE  
SEMANTIČKI SLIČNIH MEDICINSKIH SLIKA**

Rijeka, rujan 2022.

Ivana Tus

6019831100690851743

SVEUČILIŠTE U RIJECI

**TEHNIČKI FAKULTET**

Preddiplomski studij računarstva

Završni rad

**USPOREDBA GOTOVIH MODELA ZA PRONALAZENJE  
SEMANTIČKI SLIČNIH MEDICINSKIH SLIKA**

Mentor: Prof. dr. sc. Ivan Štajduhar

Rijeka, rujan 2022. .

Ivana Tus

6019831100690851743

Pod punom odgovornošću pismeno potvrđujem da je ovo moj autorski rad čiji niti jedan dio nije nastao kopiranjem ili plagiranjem tuđeg sadržaja. Prilikom izrade rada koristila sam tuđe materijale navedene u popisu literature, ali nisam kopirala niti jedan njihov dio, osim citata za koje sam navela autora i izvor, te ih jasno označila znakovima navodnika. U slučaju da se u bilo kojem trenutku dokaže suprotno, spremna sam snositi sve posljedice uključivo i poništenje javne isprave stečene dijelom i na temelju ovoga rada.

U Rijeci, rujan 2022.

---



## Sadržaj

1	Uvod . . . . .	1
2	Opis problema pretraživanja slike baziranog na sadržaju slike . . . . .	4
	2.1 Povijest pretraživanja baziranog na sadržaju slike. . . . .	5
	2.2 Tehnike izvlačenja značajki iz slika. . . . .	6
	2.2.1 Tehnike izvlačenja značajki boja iz slike . . . . .	6
	2.2.2 Tehnike izvlačenja značajki tekstura iz slike. . . . .	7
	2.2.3 Tehnike izvlačenja značajki oblika iz slike. . . . .	8
3	Gotovi modeli semantičkog pretraživanja slike po sadržaju . . . . .	9
	3.1 R-MAC algoritam na bazi ResNet modela. . . . .	9
	3.2 Duboka koalarizacija na bazi ResNet modela . . . . .	11
	3.3 Vreće konvolucijskih značajki za skalabilno pretraživanje instanci . . .	13
	3.4 Međudimenzionalna procjena za duboke konvolucijske značajke . . .	16
4	Zaključak. . . . .	18
5	Literatura. . . . .	19
6	Popis oznaka i kratica . . . . .	21
7	Sažetak i ključne riječi na hrvatskom i engleskom jeziku . . . . .	24

# 1 Uvod

Umjetna inteligencija je naziv koji se najčešće pridaje računalnim sustavima koji imaju način snalaženja u novim situacijama. U današnje doba umjetna se inteligencija koristi u razne svrhe te se u potpunosti uplela u živote običnih ljudi. Dijeli se na razne grane poput obrade prirodnoga jezika (eng. *Natural language processing*, skraćeno NPL), robotike, stručnih sustava, neuronske mreže te strojnog učenja (eng. *Machine learning*). U ovom radu orijentira se ponajviše prema posljednjoj grani, odnosno strojnom učenju. Strojno učenje definira se kao grana umjetne inteligencije koja se bavi oblikovanjem algoritama koji svoju učinkovitost poboljšavaju na temelju empirijskih podataka.[2] Tema pronalazanja semantički sličnih medicinskih slika (eng. *Content-based medical image retrieval*, skraćeno CBMIR) dio je šireg područja pretraživanja slika baziranog na sadržaju slike (eng. *Content-based image retrieval*, u daljnjem tekstu akronim CBIR) unutar kojega se koriste razni algoritmi za pronalazak slika. Ovaj rad nastavlja se na rad studenta Darija Baraća pod mentorstvom prof. dr. sc. Ivana Štajduhara s nazivom "Design and implementation of a Content-based medical image retrieval (CBMIR) system for radiology images" u kojem su objašnjeni algoritmi vreća vizualnih riječi (eng. *bag of visual words*, skraćeno BoVW), dohvat s HOG deskriptorima (histogram orijentiranih gradijenta, skraćeno HOG) te dohvat s deskriptorima izdvojenima iz konvolucijskih neuronskih mreža. Zadatak završnog rada je izdvojiti te objasniti minimalno dva dosad nespomenuta modela algoritama koji bi bili najpovoljniji za pronalazanje semantički sličnih medicinskih (radioloških) slika. Baza podataka na kojoj je baziran rad studenta Baraća sastoji se od 38.000 slika koje služe za učenje i testiranje implementacije samoga modela. Modeli pretražuju slike po 24 različita anatomska područja te u 6 različitih modaliteta, kao što je prikazano tablicama 1.1 i 1.2, pri čemu je moguće pretraživanje slike po području, po modalitetu ili oboje. Pronađene modele potrebno je usporediti i vrednovati kvalitativno i kvantitativno. Unutar podijele na anatomska područja, nalazi se podjela na više istih dijelova tijela, međutim slike u bazi podijeljene su u više područja (npr. ekstremiteti 1 i ekstremiteti 2, tablica 1.2), zbog načina označavanja medicinskog osoblja.

*Tablica 1.1 Raspodjela slika unutar baze podataka po modalitetu snimanja[1]*

<b>Modalitet snimanja</b>	<b>Dio baze podataka za treniranje algoritma</b>	<b>Dio baze podataka za testiranje algoritma</b>
RF	5123	1587
XA	5000	1961
CT	3672	1011
MR	3351	758
CR	3239	555
NM	3062	879

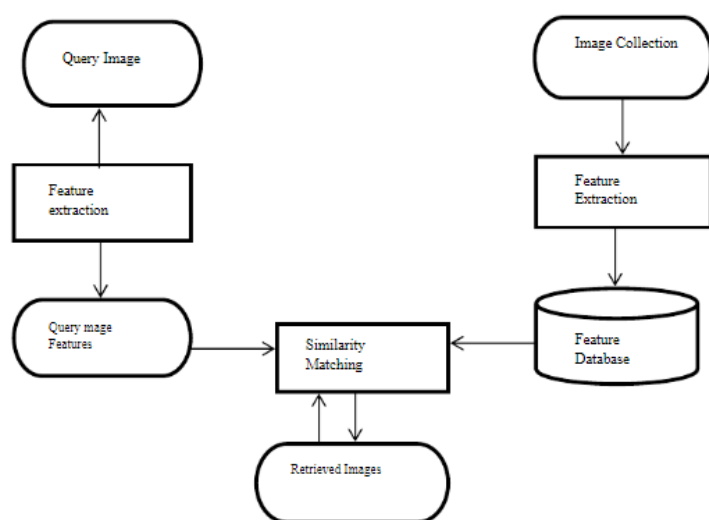


*Tablica 1.2 Raspodjela slika unutar baze podataka po anatomskom području[1]*

<b>Anatomsko područje</b>	<b>Dio baze podataka za treniranje algoritma</b>	<b>Dio baze podataka za testiranje algoritma</b>
Prsa	5123	1587
Urinarni trakt	5000	1961
Crijeva	3672	1011
Abdomen	3351	758
Glava	3239	555
Cerebrala	3062	879
Grudi	1104	432
Lumbalna kralježnica	944	65
Vratna kralježnica	877	128
Koljeno	505	55
Rame	477	68
Srce	369	69
Ruka	337	15
Zdjelica	306	118
Ekstremiteti 1	292	70
Stopalo	282	63
Vrat	260	59
Gležanj	231	16
Dlan	202	15
Kralježnica	162	22
Cijelo tijelo	77	40
Donji ekstremitet	70	5
Torakalna kralježnica	42	12
Ekstremiteti 2	16	6

## 2 Opis problema pretraživanja slike baziranog na sadržaju slike

Vid je glavno ljudsko osjetilo preko kojega primamo preko 80% podataka o okolini. Stoga ne čudi rano nastala želja za imitiranjem tog osjetila kod računala. Prvi zabilježeni zapisi sa spominjanjem računalnog vida datiraju još iz 1959. godine, no povijest pretraživanja slika može se istražiti unazad do 1979. kada je prvi put spomenuto na konferenciji Database Techniques for Pictorial Applications u Firenci.[3] Tradicionalno pretraživanje slika bazira se na ključnim riječima koje su povezane sa slikom, no to zahtijeva manualan ljudski rad označavanja svake slike s ključnim riječima što je dugotrajan posao ukoliko su u pitanju velike baze podataka te je podložno ljudskim greškama i neprepoznavanju ključnih riječi i anotacija za opis slike, te samim time smanjuje uspješnost pronalazjenja slika koje su nepotpuno ili nepravilno označene. Pretraživanje slika po sadržaju koristi isključivo informacije koje nudi sama slika poput boja, oblika, teksture i drugoga te uspoređuje sa slikama iz baze podataka za treniranje algoritma kako bi svrstao sliku u određenu kategoriju.[3][4] Pretraživanje slika po sadržaju zahtijeva dvije odvojene baze podataka: bazu podataka za učenje (treniranje) modela te bazu podataka za testiranje modela. Baza podataka za treniranje jest ona pomoću koje algoritam detektira informacije koje su ključne za svaku kategoriju slika kako bi kasnije što uspješnije prepoznao zadanu kategoriju. Ta radnja zove se izvlačenje značajki (eng. *feature extraction*) slike te je ključna za rad svakog modela prepoznavanja po sadržaju. Baza podataka za treniranje je uobičajeno veća od testne, a neke od najpoznatijih javnih baza podataka su ImageNet, Paris6k, Oxford6k, Landmarks, itd. Testna baza podataka služi za testiranje treniranog algoritma te je manja od baze podataka za treniranje. Arhitektura pretraživanja slike po sadržaju prikazana je slikom 2.1.



Slika 2.1 Generalna arhitektura modela za prepoznavanje slika po sadržaju



## 2.2 Tehnike izvlačenja značajki iz slika

Izvlačenje značajki slika je najbitnija stavka pri treniranju algoritama za pretraživanje slika po sadržaju. Izvlačenje značajki je prikupljanje svih informacija koje slika nudi kako bi se što točnije mogla svrstati u određene kategorije. Značajke slike mogu se podijeliti na značajke boja, značajke tekstura i značajke oblika unutar slike. Svaka od kategorija ima više metoda ekstrakcija značajki.

### 2.2.1 Tehnike izvlačenja značajki boja iz slike

#### Prostor boje

Najrašireniji prostor boja je RGB koji koristi različite kombinacije omjera crvene (eng. *red*, R), zelene (eng. *green*, G) i plave (eng. *blue*, B) za stvaranje ostalih boja. S obzirom da je RGB prostor boja nejednoličan, u digitalnoj se obradi slika koristi HSx prostor. HSx prostor se odnosi na HSI, HSV i HSB prostore, pri čemu H predstavlja nijansu boje (eng. *hue*), S zasićenost boje (eng. *saturation*), I intenzitet boje (eng. *intensity*), V vrijednost boje (eng. *value*) te B svjetlinu boje (eng. *brightness*). Najčešće korišteni prostor boja je HSV.[6]

#### Histogram boje

Histogram boja je tehnika izvlačenja značajki iz slike koja je spomenuta u radu "Design and implementation of a Content-based medical image retrieval (CBMIR) system for radiology images"[1] na koji se nastavlja ova rad. Histogrami boja su stupičasti grafikoni koji se oslanjaju na broj boja na slici (x-os) te broju piksela u svakoj boji (y-os). Dvije metode korištenja histograma su globalni histogram boja (eng. *global colour histogram*, GCH) te lokalni histogram boja (eng. *local colour histogram*, LCH). GCH uzima histogram cijele slike te izračunava udaljenost između histograma boja druge slike, no nedostatak mu je što ne uključuje informacije o svim regijama slike. LCH dijeli sliku na blokove te iz svakog bloka pojedinačno uzima histogram boja, a potom uspoređuje histograme boja blokova na istom mjestu na slici, zbog čega je učinkovitiji za pretraživanje slika, no računalno puno zahtjevniji od GCH-a.[6]

#### Momenti boje i geometrijski momenti

Momenti boje su mjere koje se koriste za izračunavanje sličnosti među slikama bazirano na njihovim značajkama, dok je geometrijski moment značajka koja koristi jednu vrijednost za značajku vektora. Momenti boje vide distribuciju boje na slici kao distribuciju vjerojatnosti koja je karakterizirana standardnom devijacijom i srednjom vrijednošću. Geometrijski moment, s druge

strane, producira najbolje rezultate u kombinaciji s drugim metodama ekstrakcije, no s povećanjem slike raste i količina vremena koja je potrebna za njegov izračun.[6]

### Korelogram boje

Korelogram boje je trodimenzionalna tablica indeksirana bojom i razmakom između piksela iste boje. Prikazuje kako se prostorna korelacija boje mijenja s udaljenošću te se može koristiti za usporedbu slika u bazi podataka. Da bi se napravio korelogram, boje u slici se kvantiziraju u m vrijednosti boja, što je računalno jednostavno za izračun, a opisuje globalnu distribuciju lokalnih prostornih odnosa boja.[6]

### Prosječni RGB

Prosjek boja u RGB prostoru opisan je jednadžbom

$$X = (R(avg), G(avg), B(avg))^t$$

u kojoj je:

X prosjek boja,

R(avg), G(avg) i B(avg) prosječne vrijednosti crvene, zelene i plave boje. Ova se značajka koristi u prvoj fazi pretraživanja slika filtracijom slika s većom udaljenosti u prosjeku od tražene slike.[6]

### Dominantan deskriptor boje (DCD)

Ova metoda ekstrakcije značajki boje ovisna je o histogramu boja, a odabire mali broj boja s najvišim točkama na histogramu, odnosno najdominantnije boje na slici.[6]

### Koherencijski vektor boje

Ova se metoda, kao i prethodna, temelji na histogramu boja. Dijeli histogram na koherentne komponente, čiji su pikseli prostorno povezani, i nekoherentne komponente, čiji su pikseli izolirani.[6]

## 2.2.2 Tehnike izvlačenja značajki tekstura iz slike

### Statističke metode

Prostorna distribucija sive vrijednosti izračunava se pronalaženjem lokalne značajke u svakoj točki slike te potom izvođenjem statističkih podataka iz distribucije lokalnih značajki. Primjeri tih metoda su matrica razina istodobnog pojavljivanja sive (eng. *Gray Level Co-Occurrence Matrix*,

GLCM) i Tamura značajke. GLCM je funkcija odnosa kuteva između dva piksela s odgovarajućom razinom sive te funkcija udaljenosti između njih kojom se karakterizira prostorna raspodjela sivih razina među tim pikselima. Služi za izračunavanje i razlikovanje tekstura. Tamura značajke predlažu prikaze tekstura temeljene na psihološkim studijama ljudske percepcije sa šest statističkih značajki za opis svojstava teksture: grubost, kontrast, usmjerenost, pravilnost, sličnost liniji i hrapavost.[6]

#### Pristupi temeljeni na modelu

Ova metoda izračunava proces koji je generirao teksturu. Generira nasumično polje pod pretpostavkom da je  $f(r, \omega)$ , gdje je  $r$  pozicija vektora koji reprezentira lokaciju piksela, a  $\omega$  je nasumičan parametar. Za zadanu vrijednost  $r$ ,  $f(r, \omega)$  je nasumična varijabla jer je  $\omega$  nasumičan. Za određenu teksturu  $\omega$ ,  $f(r, \omega)$  je dvodimenzionalna mreža indeksirana s  $r$  te se zove nasumično polje.[6]

#### Pristupi temeljeni na transformacijama

Najčešće transformacije na kojima se temelje metode ekstrakcije su Fourierova transformacija, diskretna valna transformacija, Ranklet transformacija te upravljiva piramida. U Fourierovoj transformaciji signal slike je prekinut u sinusne valove različitih frekvencija, međutim ne uhvati se lokacija predmeta. Diskretna valna transformacija uključuje razlaganje slike na osnovne funkcije dobivene translacijom i dilatacijom posebne funkcije te je učinkovita u analizi slike. Ranklet transformacija je dio neparametarskih, orijentacijski selektivnih i višerazlučivih značajki. Bavi se relativnim slijedom piksela umjesto vrijednostima njihovih intenziteta. Učinkovitija je od valnih i pikselnih reprezentacija slike. Upravljiva piramida generira višeskalnu i višesmjernu reprezentaciju slike. Uključuje razlaganje slike na niskopropusnom i visokopropusnom pojasu.[6]

#### 2.2.3 Tehnike izvlačenja značajki oblika iz slike

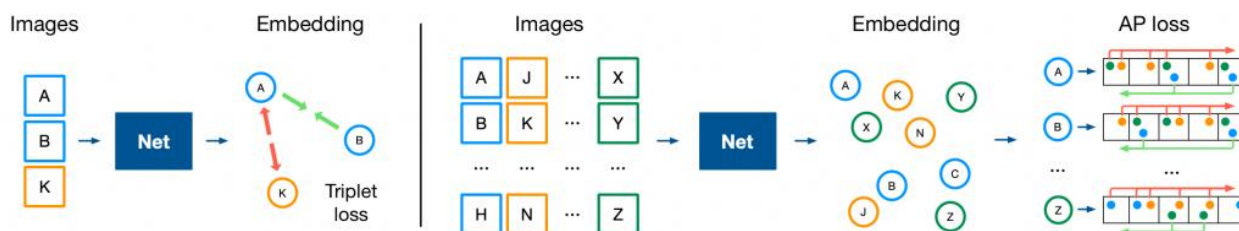
Tehnike izvlačenja značajki oblika iz slike su metoda rubova, Fourierovi deskriptori i Zernikovi momenti. Metoda rubova koristi rubove za izvlačenje informacija o obliku. Dijeli sliku u 4x4 nepreklapajuća bloka iz kojih se izvlače informacije. Tehnika Fourierovih deskriptora uključuje primjenu Fourierove transformacije na granicama oblika slike te se ti koeficijenti nazivaju Fourierovim deskriptorima. Na njih ne utječe šum. Metoda Zernikovih momenata je pogodna za složene oblike jer dopušta da se konstruira neovisni moment invarijante na svojevrijedno visok poredak.[6]

### 3 Gotovi modeli semantičkog pretraživanja slike po sadržaju

#### 3.1 R-MAC algoritam na bazi ResNet modela

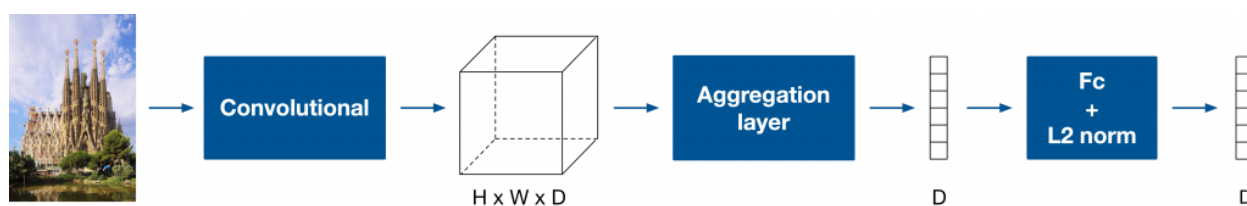
U članku "Deep Learning for Instance Retrieval: A Survey"[8] može se pronaći evaluacija gotovih modela za semantičko pretraživanje slika po sadržaju bazirano na dubokim konvolucijskim neuronskim mrežama. Među jednoprolaznim (eng. *single-pass*) modelima nalaze se regionalne maksimalne aktivacije konvolucija (eng. *Regional Maximum Activation of Convolutions*, R-MAC). R-MAC model spominje se u radovima "End-to-end Learning of Deep Visual Representations for Image Retrieval"[9] iz 2017. godine i "Learning with Average Precision: Training Image Retrieval with a Listwise Loss"[10] iz 2019. te postoji službena implementacija na GitHub-u pod repozitorijem "Deep Image Retrieval"[11]. Implementacija modela pisana je u Python-u (verzija Python 3.7.3) te PyTorch-u (verzija PyTorch 1.4) uz dodatne pakete numpy, matplotlib, tqdm i scikit-learn na bazi ResNet mreže. U članku [9] kao razlozi nezadovoljstva rezultatima dubinskih metoda za pronalaženje slika navedene su šumovite baze podataka za obučavanje, neodgovarajuća duboka arhitektura modela i neoptimalan postupak obuke mreže. Ovaj model rješava sva tri navedena problema. Kako bi se riješio problem šumovitih baza podataka, automatizirano je čišćenje dotičnih pomoću filtrirane Landmarks baze podataka koja se na samome početku zove **Landmarks-full** te se sastoji od svih slika koje se ne podudaraju s kasnije korištenim bazama Oxford6k i Paris6k. Kako bi se baza podataka očistila, pokreće se osnovna linija podudaranja slike svake klase orijentira i uspoređuje se podudaranje ključnih točaka. Nakon toga se sama podudaranja provjeravaju afinom transformacijom. Kada se prođe cijela klasa, dobije se skup uparenih rezultata te se konstruira graf s čvorovima kao slikama i rubovima kao podudaranjima. Odrežu se svi rubovi koji imaju preniski rezultat preklapanja. Na kraju se zadržavaju samo najveće povezane komponente i odbacuju ostale kako bi se osiguralo da su sve slike unutar jedne klase povezane. Takav proces čišćenja rezultira bazi koja se naziva **Landmarks-clean**. Za procjenu graničnog okvira podudaranja koristi se selektor naučenog područja (eng. *learned region of interest selector*, ROI) te se on obučava pomoću anotacija graničnog okvira koje su automatski procijenjene. Položaji provjerenih podudaranja ključnih točaka znak su predmeta interesa. Nakon izračuna i konvergacije graničnog okvira, radi se popis orijentira. Smanjenje šuma te promjena graničnog okvira mogu biti vrlo učinkovite u radu s medicinskim radiološkim slikama iz razloga što se modeli mogu u potpunosti orijentirati na bitne značajke koje pomažu prepoznavanju anatomskeg područja i/ili modaliteta.

I dok su u članku [9] u modelu predložili sijamsku arhitekturu koja kombinira tri toka s trostrukim gubitkom, u [10] predlažu gubitak srednje preciznosti kako bi se optimizirali gubitci pri rangiranju. S obzirom da se gubitci pri rangiranju uobičajeno događaju u gornjoj granici rangiranja, direktno utječu na neke metrike poput srednje prosječne preciznosti (eng. *mean average precision*, mAP). Na slici 3.1 vidi se razlika između gubitka na temelju trostrukog gubitka, koji izvodi ažuriranje gradijenata na temelju malog broja primjera, i gubitka srednje preciznosti, koji u obzir uzima velik broj slika istovremeno i time optimizira prosječnu preciznost.



Slika 3.1 Razlike između gubitka lokalnog rangiranja (lijevo: trostruki gubitak) i gubitka po listi (desno)[10]

Sama implementacija oslanja se na oba spomenuta rada te koristi konvolucijsku mrežu za izdvajanje mape svih značajki slike. Mapa je agregirana u kompaktnu prezentaciju fiksne duljine slojem globalne agregacije. Ta prezentacija je potom projicirana putem potpuno povezanog (eng. *fully-connected*, FC) sloja te je potom normalizirana pomoću L2 normalizacije, a taj proces prikazan je slikom 3.2. L2 normalizacija je tehnika kojom se modificiraju vrijednosti skupa podataka tako da u svakom retku zbroj kvadrata bude do vrijednosti 1.



Slika 3.2 Proces ekstrakcije te normalizacije mape značajki slike[11]

Unutar implementacije ponuđeno je više modela ResNet neuronske mreže. ResNet je skraćenica za rezidualnu neuronsku mrežu (eng. *residual neural network*), duboku konvolucijsku neuronsku mrežu koja se može sastojati od stotine slojeva. Unutar mreže postoji dvostruka i trostruka preskakanja slojeva s nelinearnostima (tzv. ReLu, eng. *Rectified Linear Unit*) i grupnom normalizacijom. S obzirom na preciznost modela, najbolji odabir je model ResNet101- AP-GeM. ResNet101-AP-GeM se odnosi na 101 sloj duboku ResNet mrežu s posebnim gubitkom te generaliziranim udruživanjem srednje vrijednosti (eng. *Generalized-mean pooling*, GeM). U



tablici 3.1 opisani su svi ponuđeni modeli u implementaciji pri čemu ime modela definira njegove karakteristike. U slučaju ResNet50 modela, znači da je model dubok 50 slojeva, dok je ResNet101 dubok 101 sloj. Kratica TL označava model s trostrukim gubitkom (eng. *triplet loss*) te AP označava model u kojem je korišten gubitak srednje preciznosti (eng. *average precision*). GeM i MAC se odnose na vrstu udruživanja pri čemu se GeM odnosi na već spomenuto generalizirano udruživanje srednje vrijednosti, a MAC na maksimalno udruživanje.

Tablica 3.1 Tablica ponuđenih modela unutar implementacije s postocima preciznosti u navedenim bazama podataka[11]

Model	Oxford5K	Paris6k	ROxford5K (med/hard)	RParis6K (med/hard)
Resnet101-TL-MAC	85.6	90.1	63.3/35.7	76.6/55.5
Resnet101-TL-GeM	85.7	93.4	64.5/40.9	78.8/59.2
Resnet50-AP-GeM	87.7	91.9	65.5/41.0	77.6/57.1
ResNet101-AP-GeM	89.1	93.0	67.1/42.3	80.3/60.9
ResNet101-AP-GeM-LM18	88.1	93.1	66.3/42.5	80.2/60.8

### 3.2 Duboka koalarizacija na bazi ResNet modela

Većina modela za pretraživanje slika po sadržaju obučena je na slikama u boji, što utječe na efikasnost modela pri prepoznavanju slika u sivim tonovima (tzv. crno-bijelih slika). U radu "Deep Koalarization: Image Colorization using CNNs and Inception-Resnet-v2"[13] proučava se utjecaj na učinak modela uz prethodno bojanje slika te predlaže se model kojim se bojanje može izvršiti. Ovo može biti osobito korisno u slučaju pretraživanja medicinskih radioloških slika s obzirom da su one uvijek u sivim tonovima, bez obzira na modalitet snimanja, te bi potencijalno moglo olakšati svrstavanje snimaka po anatomskim područjima. U prošlosti se bojanje slike odvijalo u skladu s povezanim referentnim slikama, a potom su se tehnike poboljšavale pomoću označavanja vrijednosti više modula svakoga piksela. Međutim, takav pristup bio je podložan greškama i različitim kontekstnim pogledima s obzirom da se oslanjao na manualan rad. Kako bi se što više smanjila mogućnost ljudske greške, počeli su se istraživati automatizirani pristupi bojanju gdje se pojavljuju konvolucijske neuronske mreže. Duboka koalarizacija predložena u [13] oslanja se na prethodno obučeni model Inception-ResNet-v2 kako bi se izvukle značajke visoke razine iz slike. Ovaj algoritam promatra sliku unutar CIE L\*a\*b trodimenzionalnog prostora boja. CIE L\*a\*b

prostor boja promatra boju kao spoj triju vrijednosti:  $L^*$  kao percipiranu svjetlinu (eng. *lightness/luminance*,  $L$ ) te vrijednosti  $a^*$  i  $b^*$  za jednu od četiri unikatne boje ljudskoga vida (crvena, zelena, plava i žuta).[14] U radu je predloženo da se, krećući od komponente svjetline, mogu procijeniti vrijednosti svih komponenti koje čine boju u CIE  $L^*a^*b^*$  prostoru. Takvo preslikavanje imalo bi jednadžbu:

$$F: X_L \rightarrow (X_a, X_b),$$

pri čemu je:

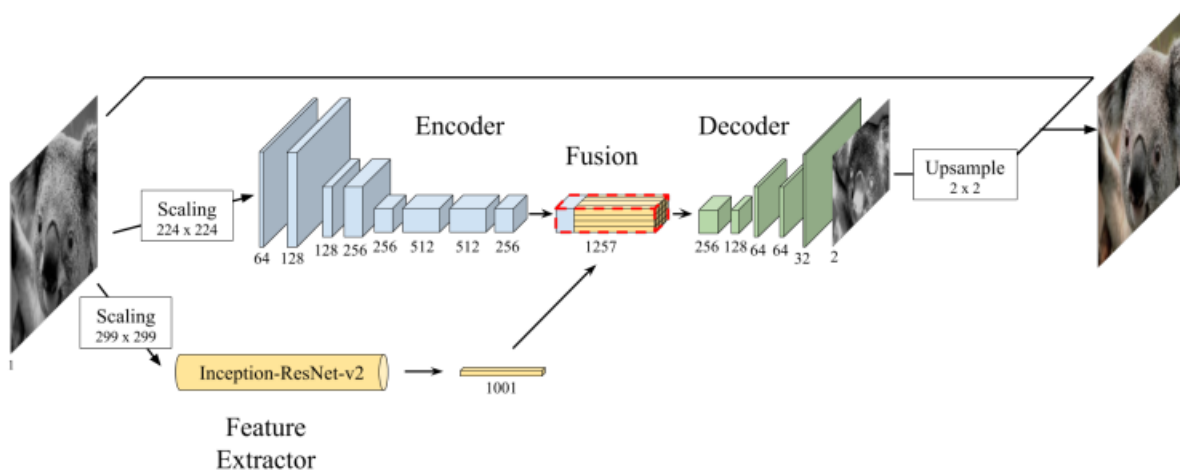
$X_L$  komponenta svjetline,

$X_a$  i  $X_b$  predstavljaju komponente  $a^*$  i  $b^*$  slike.[13]

U tom slučaju bi kombinacija tih triju komponenti

$$X = (X_L, X_a, X_b)$$

bila procijenjena obojana slika. Arhitektura modela bazirana je na spomenute tri komponente te se koristi prethodno obučenom Inception-ResNet-V2 mrežom, a podijeljena je u četiri glavna dijela: izvlačenje značajki i kodiranje, koji su zadušeni za dobavljanje značajki visokih i srednjih razina, sloj stapanja, u kojem se povezuju prethodno dobivene značajke, te dekodir koji koristi spojene značajke kako bi procijenio izlaz. Predloženi prikaz arhitekture nalazi se na slici 3.3.



Slika 3.2 Arhitektura modela za bojanje slika u sivim tonovima[13]

Kako bi se osiguralo ispravno učenje, slike se moraju obraditi u pretprocesiranju. Vrijednosti piksela sve tri komponente trebaju biti centrirane i skalirane kako bi se dobile vrijednosti unutar intervala  $[-1, 1]$ . [13] Sve upute za pripremu seta podataka nalaze se na GitHub-u službene implementacije: [github.com/baldassarreFe/deep-koalarization](https://github.com/baldassarreFe/deep-koalarization)[15]. Za treniranje modela korišteno je 54 tisuće slika, dok je 6 tisuća sačuvano za validaciju. Rezultati pokazuju da elementi

koji su se pojavljivali u većem dijelu slika pri treniranju, generiraju izvrsne rezultate. Primjer bojanja slike ovim algoritmom nalazi se na slici 3.3.



Slika 3.3 Usporedba rezultata dobivenih korištenjem duboke koalarizacije za bojanje (srednji stupac) sa slikom u sivim tonovima (lijevi stupac) i stvarne slike (desni stupac)[13]

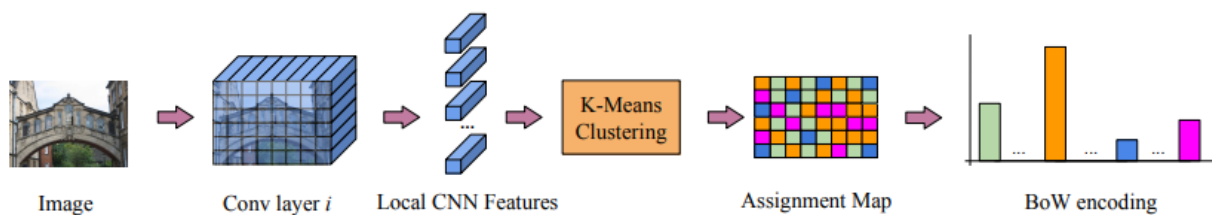
### 3.3 Vreće konvolucijskih značajki za skalabilno pretraživanje instanci

Algoritam vreće vizualnih riječi spomenut je u Uvodu ovoga rada kao algoritam obrađen u radu koji je prethodio ovome. Međutim, taj algoritam prethodio je daljnjim istraživanjima te je temelj za neke nove modele koji u priču dodaju i konvolucijske neuronske mreže. Jedan od tih modela je i model vreća konvolucijskih značajki za skalabilno pretraživanje instanci objašnjen u radu "Bags of Local Convolutional Features for Scalable Instance Search". Rad je predstavljen 2016. godine na Međunarodnoj konferenciji o multimedijском pronalaženju (eng. *International Conference on Multimedia Retrieval*, ICMR), organiziranoj od strane Udruge za računalne strojeve (eng. *Association for Computer Machinery*, ACM). Model kombinira deskriptore konvolucijskih

neuronskih mreža sa shemom kodiranja vreće riječi (eng. *Bag of Words*, BoW) te je predstavio tri noviteta na području pretraživanja slika po sadržaju:

- prijedlog oskudnog vizualnog deskriptora temeljenog na vreći lokalnih konvolucijskih značajki (eng. *Bag of Local Convolutional Features*, BLCF) što omogućuje brzo pronalaženje slike pomoću obrnutog indeksa[16]
- predstavljanje mape dodjele kao kompaktnog prikaza slike koji preslikava piksele u slike na njihove odgovarajuće vizualne riječi što dopušta brzo sastavljanje BoW deskriptora za bilo koje područje slike[16]
- iskorištavanje svojstva skalabilnosti mape dodjele za izvođenje lokalne analize više regija slike za ponovno rangiranje, nakon čega slijedi faza proširenja upita korištenjem dobivenih lokalizacija objekata.[16]

Iako se i u prethodnim radovima i modelima koristila konvolucijska neuronska mreža za izvlačenje značajki, uobičajeno se to izvlačenje odvijalo na globalnoj razini slike. Time bi neke lokalne značajke ostale zanemarene, a samo pretraživanje bi rezultiralo s manjom točnošću. Taj problem pokušao se riješiti koristeći blokove slika iz kojih bi CNN izvukao značajke (za CNN globalne, no s obzirom da se koristio blokovi lokalne za taj dio ukupne slike) te ih se naposljetku kombiniralo sa značajkama ostalih blokova putem prosječnog udruživanja, maksimalnog udruživanja ili vektora lokalno agregiranih deskriptora (eng. *Vector of Locally Aggregated Descriptor*, VLAD). Zbog korištenja velikog broja blokova, ovakav način pretraživanja slika zahtijeva mnogo vremena za indeksiranje i izvlačenje značajki. BCLF model predlaže jedan cjevovod kroz CNN u kojemu se aktivacije različitih nizova neurona u mapama značajki tretiraju kao lokalne značajke u konvolucijskom sloju te su potom kodirani pomoću BoW-a. BoW kodiranje lokalnih konvolucijskih značajki u vektor korišteno je iz nekoliko razloga: BoW kodiranje proizvode oskudne visokodimenzionalne kodove koji se potom pohranjuju u invertirane indekse te su korisni za brzo pronalaženje, a prikazi bazirani na BoW-u kompaktniji su te su brži za računanje i lakši za protumačiti.[16] BoW model traži izradu vizualnog šifrnika koji mapira vektore u njihov najbliži centroid, a potom se svaka lokalna značajka izvučena CNN-om dodjeljuje svojoj najbližoj vizualnoj riječi u šifrniku. Šifrnici je dvodimenzionalna mapa dodjele koja povezuje svaki piksel slike s njegovom vizualnom riječi te time omogućuje brzo generiranje BoW vektora (za cijelu sliku i njezine dijelove).[16] Sama arhitektura cjevovoda ovoga modela sastoji se od tri dijela: početna pretraga i rangiranje, ponovno prostorno rangiranje te proširenje upita. Arhitektura samog cjevovoda nalazi se na slici 3.4.



Slika 3.4 Arhitektura cjevovoda BLCF modela

Početna pretraga i rangiranje odvija se na dvije razine: globalno pretraživanje, u kojem je BoW vektor upita izrađen vizualnim riječima svih lokalnih značajki konvolucijskog sloja za sliku upita, te lokalno pretraživanje, u kojem se BoW vektor upita sastoji od samo jedne vizualne riječi lokalnih značajki koje spadaju unutar granica upita. Početni se poredak tada izračunava pomoću sličnosti kosinusa između BoW vektora slike upita i BoW vektora cijelih slika u bazi podataka. Koristi se obrnuti indeks temeljen na oskudnoj matrici i množenje oskudne matrice temeljeno na grafičkoj procesnoj jedinici (eng. *graphics processing unit*, GPU) za brzu pretragu baze podataka. Na najbližih sto slika u početnoj pretrazi primjenjuje se lokalno ponovno rangiranje pomoću pristupa kliznog prozora (eng. *sliding window approach*). To znači da se odaberu sve kombinacije prozora visine  $h$  i širine  $w$  koje su omjeri visine  $H$  i širine  $W$  mape dodjele te se pretražuju lokalne značajke s preklapanjem od 50% između prozora. Potom se rangiraju pojedini prozori formulom

$$score_w = \frac{\min(AR_w, AR_q)}{\max(AR_w, AR_q)}$$

pri čemu je:

$$AR_q = W_q/H_q \text{ omjer graničnog okvira upita,}$$

$$AR_w = W_w/H_w \text{ omjer širine i visine prozora.}$$

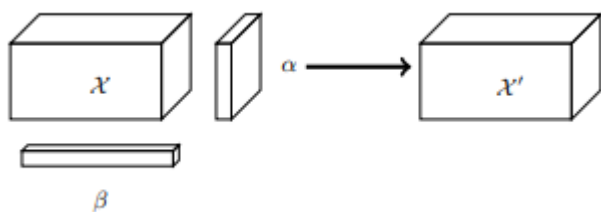
Svi se prozori čiji je rezultat ispod praga zanemaruju. Za preostale prozore konstruira se BoW vektor i uspoređuje s BoW vektorom upita pomoću sličnosti kosinusa te se bira prozor s najvećom sličnosti kosinusa. Potom se odvija proširenje upita na globalnoj razini, na kojoj se BoW vektori  $N$  slika na vrhu poretka udružuju s BoW vektorom upita kako bi nastala nova reprezentacija upita, i na lokalnoj razini, na kojoj se lokacije u koraku ponovnog rangiranja koriste za konstrukciju BoW deskriptora samo od područja interesa  $N$  slika na vrhu poretka te se potom ti vektori udružuju s BoW vektorom graničnog okvira upita kako bi se izvršilo drugo pretraživanje.[16] Pri izvlačenju značajki koristi se prethodno obučena VGG16 neuronska mreža te se značajke izvlače iz barem triju konvolucijskih slojeva. Također, koristi se L2 normalizacija lokalnih značajki objašnjena u potpoglavlju 3.1.

### 3.4 Međudimenzionalna procjena za duboke konvolucijske značajke

Kad se govori o izvlačenju značajki iz slike, one se uvijek dijele na lokalne i globalne značajke. I dok globalne značajke opisuju sliku u cjelini, lokalne značajke mogu dati više informacija o bitnim detaljima koje mogu uvelike poboljšati točnost pretrage slika. Stoga je bitno u obzir uvesti i globalne i lokalne značajke kao sveukupne reprezentacije slike. U radu "Cross-dimensional Weighting for Aggregated Deep Convolutional Features" predstavlja se jedan od načina agregacije svih značajki za bolju globalnu reprezentaciju slike te potom pretraživanje slika pomoću upita s dobivenom reprezentacijom. Model naziva međudimenzionalna procjena za duboke konvolucijske značajke (eng. *Cross-dimensional Weighting for Aggregated Deep Convolutional Features*, CroW) temelji se na neparametarskoj shemi ponderiranja i agregacije konvolucijskih značajki izvučenih pomoću VGG16 neuronske mreže. Rad prati i službena dokumentacija objavljena na GitHub-u pod istoimenim repozitorijem[19]. Sam algoritam modela odvija se u 7 koraka: prostorno-lokalno udruživanje, izračunavanje lokalnih, a potom i kanalnih težinskih vektora, agregacija zbroja težina te potom dvije normalizacije dobivenog vektora i smanjenje njegove dimenzionalnosti. Prvo se izvodi prostorno-lokalno udruživanje susjedstva kojime se dobije trodimenzionalan tenzor aktivnosti. Potom se izračunavaju faktori prostorne težine i faktori težine kanala. Faktori prostorne težine označavaju se s  $\beta_k$  gdje je  $k$  oznaka kanala, dok se faktori prostorne težine označavaju s matricom  $\alpha_{ij}$  u kojoj su  $i$  i  $j$  oznake lokacije. Time za tenzor  $X'_{kij}$  vrijedi:

$$X'_{kij} = \alpha_{ij}\beta_k X_{kij}.$$

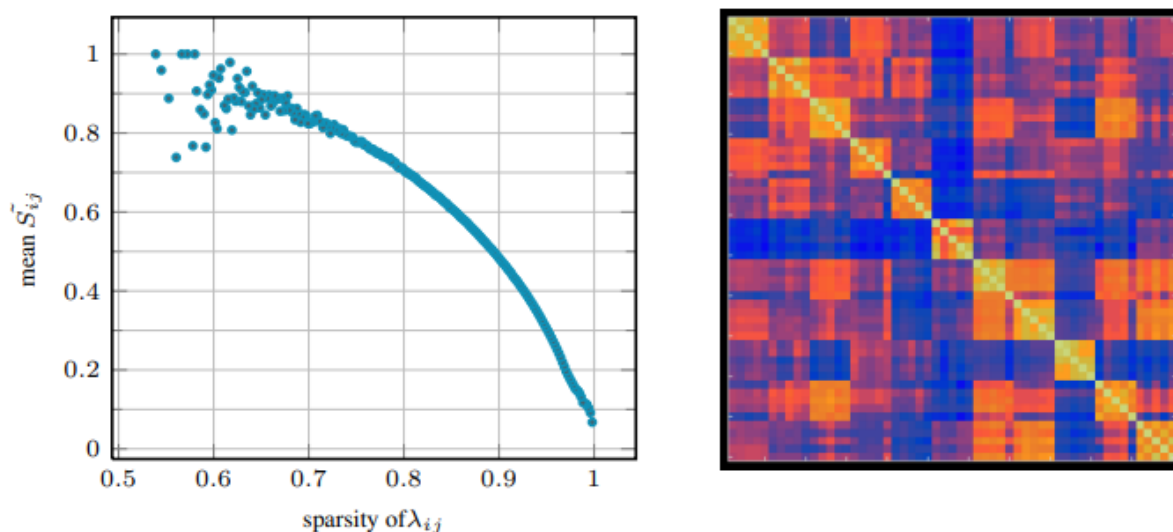
To je grafički prikazano slikom 3.5.



Slika 3.5 Prikaz tenzora  $X'$  te njegove ovisnosti od tenzoru  $X$ , težinskoj matrici  $\alpha$  i težinskom vektoru  $\beta$

Nakon izračunavanja potrebnih težina, izvodi se agregacija ponderiranog zbroja punog tenzora u jedan vektor koji se potom normalizira te mu se smanjuje dimenzionalnost. To se najčešće radi analizom glavnih komponenti (eng. *principal component analysis*, PCA), no mogu se koristiti i

druga skaliranja. Nakon skaliranja izvodi se druga i posljednja normalizacija dobivenog vektora. Što se tiče ponderiranja značajki, uvode se dvije sheme: agregacija odgovora za prostorno ponderiranje i ponderiranje kanala osjetljivo na rijetkost. Agregacija odgovora za prostorno ponderiranje temelji se na normalizaciji ukupnog odgovora na svim kanalima, a takvo prostorno ponderiranje pojačava značajke na lokacijama s istaknutim vizualnim sadržajem i smanjuje pondere na neistaknutim lokacijama. Na slici 3.6 na grafu s lijeve strane prikazan je odnos između prostornih težina  $S_{ij}$  i rijetkosti odziva kanala  $\lambda^{ij}$ . Ovakvo prostorno ponderiranje ima tendenciju povećavanja lokacija za koje su aktivni višestruki kanali te to sugerira neparametarski i računalno jeftin način favoriziranja prostorne lokacije. Ponderiranje kanala osjetljivo na rijetkost metoda je za izvođenje težine signala na temelju rijetkosti mapa značajki. Na slici 3.6 na grafu s desne strane nalazi se vizualizacija korelacije para vektora razrijeđenosti kanala  $\Xi \in R^K$ , gdje je  $\Xi$  razrijeđenost kanala i  $K$  oznaka broja kanala, za slike u setu upita Paris6k bazi podataka. Razrijeđenosti kanala visoko su povezane za slike istog orijentira, a manje za slike različitih orijentira.



*Slika 3.6 S lijeve strane se nalazi graf odnosa prostornih težina u odnosu na rijetkost odziva kanala*

*S desne strane nalazi se graf korelacije razrijeđenosti kanala skupa upita Paris6k skupa podataka, poredanih prema klasi orijentira u obje dimenzije [18]*

Ovakav način ponderiranja i agregacije daje vrhunske rezultate uz minimalne računalne troškove.

## 4 Zaključak

Pretraživanje slika po sadržaju budućnost je pretraživanja slika te je sam razvoj tog područja umjetne inteligencije tek započeo. Proteklih godina pokazalo se da modeli bazirani na konvolucijskim neuronskim mrežama daju vrlo dobre rezultate, ali koji model je najbolji, ne može se sa sigurnošću tvrditi. Model duboke koalarizacije pokazuje visoku preciznost na svim bazama podataka, a zbog mogućnosti bojanja slika u sivim tonovima, može se iskoristiti za lakše prepoznavanje anatomskih područja na radiološkim slikama. Također, za napomenuti je da je to jedini višeprolazni model obrađen u ovom radu, odnosno kod drugih modela slika samo jednom prolazi kroz „cjevovod“ za izvlačenje značajki i pretraživanje slika, dok kroz „cjevovod“ modela duboke koalarizacije prolazi više puta, što može poboljšati i točnost određivanja modaliteta snimanja radioloških slika. Međutim u obzir treba dovesti i računalnu zahtjevnost, a jednoprolazni modeli imaju prednost na tom području. S obzirom da u jednoprolaznim modelima slika kroz proces izvlačenja značajki i pretraživanja slika prolazi samo jednom, samo vrijeme obučavanja modela na određenoj bazi podataka može biti kraće od višeprolaznih modela. Usporedba opisanih modela na različitim referentnim bazama podataka može se naći u tablici 4.1.

*Tablica 4.1 Usporedba spomenutih modela na nekoliko referentnih baza podataka u postocima točnosti[8]*

<b>Tip</b>	<b>Metoda</b>	<b>Holidays</b>	<b>UKB</b>	<b>Oxford6k</b> (+100k)	<b>Paris6k</b> (+100k)
jednoprolazni	R-MAC	-	-	66.9 (61.6)	- (75.7)
višeprolazni	Duboka koalarizacija	84.3	91.1 (mAP)	68.0 (-)	79.5 (-)
jednoprolazni	BLCF	-	-	73.9 (59.3)	82.0 (64.8)
jednoprolazni	CroW	85.1	-	68.4 (63.7)	76.5 (69.1)



## 5 Literatura

[1] Barać, D., Štajduhar, I.: "Design and implementation of a Content-based medical image retrieval (CBMIR) system for radiology images", Tehnički fakultet Sveučilišta u Rijeci, Rijeka, 2021.

[2] Strojno učenje 1, s Interneta:

<https://www.fer.unizg.hr/predmet/struce1> (pristupljeno 15. svibnja 2022.)

[3] Long, F., Zhang, H., Feng, D.D.: "Fundamentals of Content-Based Image Retrieval", Springer-Verlag Berlin Heidelberg, Berlin, 2003.

[4] Wikipedija - Content-based image retrieval, s interneta: [https://en.wikipedia.org/wiki/Content-based\\_image\\_retrieval](https://en.wikipedia.org/wiki/Content-based_image_retrieval) (pristupljeno 1. kolovoza 2022.)

[5] Jain, S., Pulaparthy, K., Fulara, C.: "Content based image retrieval", Maharaja Agrasen Institute of Technology, GGSIP Univeristy New Delhi, New Delhi, 2015.

[6] Olaleke, J., Adetunmbi, A., Ojokoh, B., Olaronke, I.: "An Appraisal of Content-Based Image Retrieval (CBIR) Methods", ResearchGate, Nigeria, 2019.

[7] Krizhevsky, A., Sutskever, I., Hinton, G.: "ImageNet Classification with Deep Convolutional Neural Networks", University of Toronto, Toronto, 2012.

[8] Wei, C., Yu, L., Weiping, W., Bakker, E., Georgiou, T., Fieguth, P., Li, L., Lew, M.: "Deep Learning for Instance Retrieval: A Survey"

[9] Gordo, A., Almazan, J., Revaud, J., Larlus, D.: "End-to-end Learning of Deep Visual Representations for Image Retrieval", International Journal of Computer Vision (IJCV), 2017.

[10] Revaud, J., Almazan, J., Sampaio de Rezende, R., de Souza, C. : "Learning with Average Precision: Training Image Retrieval with a Listwise Loss", International Conference on Computer Vision (ICCV), 2019.

[11] Github - Deep image retrieval, s interneta:

<https://github.com/naver/deep-image-retrieval> (pristupljeno 28.8.2022.)

[12] Wikipedija - Residual neural network, s interneta: [https://en.wikipedia.org/wiki/Residual\\_neural\\_network](https://en.wikipedia.org/wiki/Residual_neural_network) (pristupljeno 28.8.2022.)

- [13] Baldassarre, F., Gonzalez Morin, D., Rodes-Guirao, L.: "Deep Koalarization: Image Colorization using CNNs and Inception-Resnet-v2", KTH Royal Institute of Technology, Stockholm, 2017.
- [14] Wikipedija - CIELAB color space, s interneta: [https://en.wikipedia.org/wiki/CIELAB\\_color\\_space](https://en.wikipedia.org/wiki/CIELAB_color_space) (pristupljeno 8.9.2022.)
- [15] Github - Deep koalarization, s interneta: <https://github.com/baldassarreFe/deep-koalarization> (pristupljeno 8.9.2022.)
- [16] Mohedano, E., McGuinness, K., O'Connor, N., Salvador, A., Marqu'es, F., Gir'o-i-Nieto, X.: "Bags of Local Convolutional Features for Scalable Instance Search", ACM International Conference on Multimedia Retrieval (ICMR), New York, 2016.
- [17] Github - Bags of Local Convolutional Features for Scalable Instance Search, s interneta: <https://github.com/imatge-upc/retrieval-2016-icmr> (pristupljeno 9.9.2022.)
- [18] Kalantidis, Y., Mellina, C., Osindero, S.: "Cross-dimensional Weighting for Aggregated Deep Convolutional Features", Computer Vision and Machine Learning Group Flickr, Yahoo, 2016.
- [19] Github - Cross-dimensional Weighting (CroW) aggregation, s interneta: <https://github.com/YahooArchive/crow> (pristupljeno 10.9.2022.)

## 6 Popis oznaka i kratica

Kratica	Puni naziv na stranom jeziku	Tumačenje na hrvatskom jeziku
ACM	eng. <i>Association for Computer Machinery</i>	Udruga za računalne strojeve
AP	eng. <i>Average precision</i>	Prosječna preciznost
avg	eng. <i>Average</i>	Prosjek
BLCF	eng. <i>Bag of Local Convolutional Features</i>	Vreća lokalnih konvolucijskih značajki
BoVW	eng. <i>Bag of visual words</i>	Vreća vizualnih riječi
BoW	eng. <i>Bag of Words</i>	Vreća riječi
CBIR	eng. <i>Content-based image retrieval</i>	Pronalazak semantički sličnih slika
CBMIR	eng. <i>Content-based medical image retrieval</i>	Pronalazak semantički sličnih medicinskih slika
CIE	eng. <i>International Commission on Illumination</i>	Međunarodna komisija za rasvjetu
CNN	eng. <i>Convolutional neural network</i>	Konvolucijska neuronska mreža
CR	eng. <i>Computed radiography</i>	Kompjuterizirana radiografija
CroW	eng. <i>Cross-dimensional Weighting</i>	Međudimenzionalno ponderiranje
CT	eng. <i>Computed tomography</i>	Kompjuterizirana tomografija
DCD	eng. <i>Dominant colour descriptor</i>	Dominantan deskriptor
FC	eng. <i>Fully-connected</i>	Potpuno povezano
HOG	eng. <i>Histogram of Oriented Gradients</i>	Histogram orijentiranih gradijenata
HSB	eng. <i>Hue, saturation, brightness</i>	Nijansa, zasićenost, svjetlina

<b>Kratica</b>	<b>Puni naziv na stranom jeziku</b>	<b>Tumačenje na hrvatskom jeziku</b>
HSI	eng. <i>Hue, saturation, intensity</i>	Nijansa, zasićenost, intenzitet
HSV	eng. <i>Hue, saturation, value</i>	Nijansa, zasićenost, vrijednost
GCH	eng. <i>Global colour histogram</i>	Globalni histogram boja
GeM	eng. <i>Generalized-mean pooling</i>	Generalizirano udruživanje srednje vrijednosti
GLCM	eng. <i>Gray Level Co-Occurrence Matrix</i>	Matrica razina istodobnog pojavljivanja sive
GPU	eng. <i>Graphics processing unit</i>	Grafička procesna jedinica
ICMR	eng. <i>International Conference on Multimedia Retrieval</i>	Međunarodna konferencija o multimedijском pronalaženju
ILSVRC	eng. <i>ImageNet Large Scale Visual Recognition Challenge</i>	ImageNet-ov izazov vizualnog prepoznavanja na velikoj skali
L	eng. <i>Lightness/Luminance</i>	Svjetlina
LCH	eng. <i>Local colour histogram</i>	Lokalni histogram boja
MAC	eng. <i>Max-pooling</i>	Maksimalno udruživanje
mAP	eng. <i>Mean average precision</i>	Srednja prosječna vrijednost
MR	eng. <i>Magnetic resonance</i>	Magnetska rezonanca
NM	eng. <i>Nuclear medicine</i>	Nuklearna medicina
NPL	eng. <i>Natural language processing</i>	Prirodna obrada jezika
PCA	eng. <i>Principal component analysis</i>	Analiza glavnih komponenti
ReLu	eng. <i>Rectified Linear Unit</i>	Ispravljena linearna jedinica
ResNet	eng. <i>Residual neural network</i>	Rezidualna neuronska mreža
RF	eng. <i>Radio frequency</i>	Radijska frekvencija

<b>Kratica</b>	<b>Puni naziv na stranom jeziku</b>	<b>Tumačenje na hrvatskom jeziku</b>
RGB	eng. <i>Red, green, blue</i>	Crvena, zelena, plava
R-MAC	eng. <i>Regional Maximum Activation of Convolutions</i>	Regionalna maksimalna aktivacija konvolucija
ROI	eng. <i>Region of interest</i>	Područje interesa
SFM	eng. <i>Structure from motion</i>	Struktura iz kretanja
TL	eng. <i>Triplet loss</i>	Trostruki gubitak
VLAD	eng. <i>Vector of Locally Aggregated Descriptor</i>	Vektor lokalnih agregiranih deskriptora
XA	eng. <i>X Ray Angiography</i>	Rendgenska angiografija

## 7 Sažetak

U ovome radu istraženo je područje umjetne inteligencije pod nazivom računalni vid, konkretno pretraživanje medicinskih slika po sadržaju. Rad prolazi kroz povijest pretraživanja slika i objašnjava njegovu važnost i razvoj. Objasnjene su značajke slika koje se izvlače kako bi se koristile za pronalazak slike te se spominju različiti načini njihove ekstrakcije. Opisana su četiri različita modela pretraživanja slike po sadržaju i njihove metode izvlačenja značajki i pronalaska slike. Također, modeli su smješteni u područje pretraživanja medicinskih slika i dana je usporedba između njih.

**Ključne riječi:** pretraživanje slika po sadržaju, konvolucijske neuronske mreže, izvlačenje značajki

### **Abstract**

In this paper the field of artificial intelligence called computer vision, in particular content based image retrieval, is explored. The paper goes through history of image retrieval and describes its importance and development. The image features that are extracted in order to retrieve images are explained and different techniques for their extraction are mentioned. Four different off-the-shelf models for content based image retrieval are mentioned and their ways of feature extraction are explained. The models are placed in the field of medical image retrieval and a comparison between them is given.

**Keywords:** content based image retrieval, convolutional neural networks, feature extraction