

**ALGORITMI DETEKCIJE OBJEKATA PRI DIGITALNOJ OBRADI FOTOGRAFIJE****OBJECT DETECTION ALGORITHMS FOR DIGITAL IMAGE PROCESSING**Katarina Gavrilov, *Fakultet tehničkih nauka, Novi Sad***Oblast – INŽENJERSTVO INFORMACIONIH SISTEMA**

**Kratak sadržaj** – U radu je opisan pojam računarske vizije i dati su primeri njene upotrebe. Objašnjeni su osnovni koncepti digitalne obrade fotografije i dat je prikaz vrsta algoritama detekcije objekata pri digitalnoj obradi fotografije, koji obuhvata uvid u polja njihove upotrebe.

**Gljučne reči:** Računarska vizija, konvolucione neuralne mreže, algoritmi detekcije objekata, digitalna obrada fotografije

**Abstract** – The paper describes the concept of computer vision and gives examples of its use. The basic concepts of digital photo processing are explained and algorithms for detecting objects in digital photo processing are presented, which include insight into the fields of their use.

**Keywords:** Computer Vision, Convolutional Neural Networks, object detection algorithms, digital image processing

**1. UVOD**

Osnovni cilj ovog rada je izvršiti analizu postojećih algoritama za detekciju objekata pri digitalnog obradi fotografije. Zadaci za realizaciju cilja su:

- identifikacija algoritama za detekciju objekata i
- analiza algoritama.

Struktura rada organizovana je po sledećim poglavljima. U prvom poglavlju predstavljen je pojam računarske vizije, uz opis primarnih grana njene primene. Naredno poglavlje sadrži osnovne koncepte digitalne obrade fotografije. U sledećem poglavlju predstavljeno je objašnjenje algoritma detekcije objekata, kao i opis njegove upotrebe i postojećih vrsta. Poslednje poglavlje je zaključak, u kom su predstavljena razmatranja i mogućnosti za dalji rad.

**2. POJAM RAČUNARSKE VIZIJE**

Iako se računarska vizija (CV, engl. Computer Vision) posmatra kao relativno mlado i neistraženo naučno polje, razumevanje vizuelnih podataka i pokušavanje razvijanja načina na osnovu kojih mašine vide, datira još iz 1959. godine.

Te godine neurofiziolozi su dali niz fotografija na uvid mački, kako bi pokušali da razumeju način povezivanja

**NAPOMENA:**

**Ovaj rad proistekao je iz master rada čiji mentor je bio prof. dr Srđan Sladojević.**

odgovora u njenom mozgu. Tim eksperimentom su otkrili da je prva reakcija bila na istaknute ivice ili linije, što bi sa naučnog aspekta značilo da obrada fotografija počinje jednostavnim oblicima [1]. U istom periodu. došlo je i do razvijanja prve tehnologije skeniranja računarskih fotografija, koja je omogućila digitalizaciju fotografija. Još jedno bitno dostignuće desilo se 1963. godine, kada je izvršena uspešna transformacija dvodimenzionalnih fotografija u trodimenzionalne oblike [1].

Tehnologija optičkog prepoznavanja znakova (OCR, engl. *Optical character recognition*), uvedena je 1974. godine, a nakon nje i inteligentno prepoznavanje znakova (ICR, engl. *Intelligent character recognition*). Primena OCR-a i ICR-a je u obradi dokumenata i računa, prepoznavanja tablica vozila, mobilnom plaćanju i slično [1]. U periodu 1980-1982. godine, došlo je do novih istraživanja. Naučnik David Mar uveo je algoritme za detekciju ivica, uglova, krivih i sličnih oblika [2], dok je naučnik Kunihiko Fukushima razvio mrežu ćelija koje mogu prepoznati obrasce.

Računarska vizija predstavlja deo polja veštačke inteligencije (AI, engl. *Artificial Intelligence*). Osnovni cilj računarske vizije jeste izvlačenje korisnih informacija iz fotografija. Omogućava računarima da pristupe značajnim informacijama iz digitalnih fotografija, video zapisa i ostalih vizuelnih ulaznih podataka – na osnovu kojih računari mogu da izvrše određene zadatke ili da dobiju preporuke za dalji rad na osnovu dobijenih informacija. Računarima veštačka inteligencija omogućava da razmišljaju, ali računarska vizija im omogućava posmatranje i razumevanje. Da bi računarska vizija postigla što bolje rezultate, potrebno je mnogo podataka. Određena mašina iznova i iznova vrši analizu podataka, sve dok ne uoči razlike i ne prepozna fotografije. Za uspešno postizanje rezultata, u okviru računarske vizije, koriste se dve tehnologije: duboko učenje (DL, engl. *Deep Learning*) i konvoluciona neuronska mreža (CNN, engl. *Convolutional neural network*).

Aplikacije za računarsku viziju postaju sveprisutne u našem svakodnevnom životu. Možemo pronaći objekat ili lice, razumeti kretanje i obrasce u video zapisu i povećati ili smanjiti veličinu, osvetljenost ili oštrinu fotografije. Tačnost aplikacija za računarsku viziju zavisi od toga koliko su fotografije ili video zapisi interpretirani.

**3. OSNOVNI KONCEPTI DIGITALNE OBRADE FOTOGRAFIJE**

Fotografija je prostorni prikaz dvodimenzionalne ili trodimenzionalne scene. Predstavlja se kao niz ili matrični

piksel raspoređen u kolone i redove. Fotografija takođe podrazumeva dvodimenzionalni niz posebno raspoređen u redove i kolone. Digitalna fotografija se sastoji od konačnog broja elemenata, od kojih svaki ima lokaciju i vrednost. Ovi elementi se nazivaju element fotografije i pikseli. Pikseli se koriste za označavanje elementa digitalne fotografije [3].

Digitalna fotografija u sivim tonovima je predstavljena u računaru matricom piksela. Svaki piksel takve fotografije predstavlja jedan element matrice - jedan ceo broj iz skupa.

Numeričke vrednosti u prezentaciji piksela ravnomerno se menjaju od 0 do 255. Vrednost broja 0 prikazuje crne piksele, dok se beli pikseli predstavljaju sa vrednošću broja 255 [3].

Tipovi fotografija koji se najčešće pojavljuju su:

1. Binarne fotografije. Kod binarnih fotografija, svaki piksel se sastoji od jednog bita, gde taj bit može da uzme vrednost 0 (crna boja) ili 1 (bela boja).
2. Fotografije sa sivim tonovima. Fotografije sa sivim tonovima sadrže piksele koji se sastoje od 8 bitova. To podrazumeva da svaki piksel može da preuzme jednu vrednost sive boje u rasponu od 0 do 255. U ovom slučaju, crna boja je, kao i kod binarnih fotografija predstavljena brojem 0, dok je bela boja predstavljena brojem 255.
3. Fotografije u boji (RGB fotografije). Pikseli u ovim fotografijama su predstavljeni kao kombinacija 3 boje (crvene, zelene i plave) i sve druge boje se dobijaju mešanjem ove tri boje u određenim količinama. Pikseli u ovim fotografijama se sastoje iz trodimenzionalnog vektora, gde svaka od 3 vrednosti predstavlja prisutnost jedne od tri boje u tom pikselu. Takođe, važno je napomenuti da se vrednosti u vektoru predstavljaju sa 8 bitova (od 0 do 255).

Digitalna obrada fotografija predstavlja manipulaciju digitalnim podacima uz pomoć računara kako bi se kreirale digitalne mape, unutar kojih su ključne informacije izdvojene i istaknute. Digitalna obrada fotografija omogućava poboljšavanje karakteristika fotografije koje umanjuju detalje koji su nebitni za datu aplikaciju, a zatim iz poboljšane fotografije izdvajaju korisne informacije o sceni.

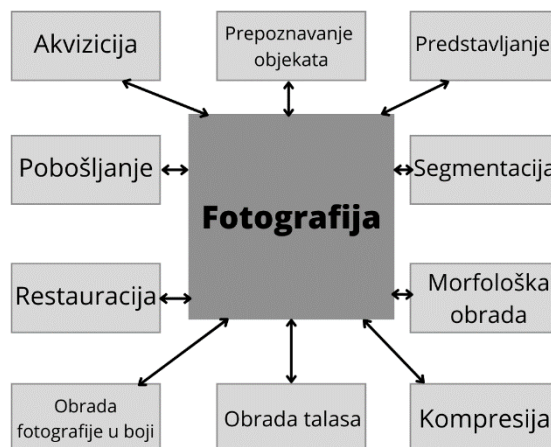
Fotografija se može definisati kao dvodimenzionalna funkcija,  $f(x, y)$ , gde su  $x$  i  $y$  prostorne (ravni) koordinate, a amplituda  $f$  na bilo kom paru koordinata  $(x, y)$  se naziva intenzitet ili nivo sive fotografije u toj tački. Kada su  $x$  i  $y$  amplitudne vrednosti  $f$  sve konačne, diskretne veličine, fotografiju nazivamo digitalnom fotografijom.

Područje digitalne obrade fotografija odnosi se na obradu digitalnih fotografija pomoću digitalnog računara [4].

### 3.1. KORACI PRI DIGITALNOJ OBRADI FOTOGRAFIJA

Organizacija koraka koji se koriste pri digitalnoj obradi fotografije je predstavljena na slici 1. Korak koji će se

koristiti zavisi isključivo od namene i cilja obrade fotografije [5].



SLIKA 1. ORGANIZACIJA KORAKA PRI DIGITALNOJ OBRADI FOTOGRAFIJE

Kategorije koje se podrazumevaju kao koraci pri digitalnoj obradi fotografija su:

1. Akvizicija – uključuje prethodnu obradu koja obuhvata:
  - a) skaliranje i
  - b) konverziju boja (iz RGB prostora boja u sivu nijansu ili obrnuto);
2. Pobošljanje fotografije – jedan od najjednostavnijih načina obrade fotografije. Koristi se za otkrivanje zatamnenih detalja i isticanje određenih karakteristika na fotografiji, upotrebom promene osvetljenosti, kontrasta, i slično;
3. Restauracija fotografije – proces poboljšavanja izgleda fotografije. Tehnike restauracije se obično zasnivaju na matematičkim modelima degradacije fotografije;
4. Obrada fotografije u boji – uključuje rad sa pseudobojom i modelima u boji za obradu fotografija u boji koji se primenjuju pri digitalnoj obradi fotografija;
5. Obrada talasa i obrada više rezolucija – omogućavaju temelj za predstavljanje fotografija u različitim stepenima rezolucije. Fotografije se dele na manje regione zbog kompresije podataka i piramidalnog predstavljanja;
6. Kompresija fotografije – tehnike ove kategorije smanjuju memoriju koja je potrebna za otpremanje fotografije ili za njen prenos. Kompresija najčešće podrazumeva promenu veličine fotografije ili njene rezolucije;
7. Morfološka obrada – proces izdvajanja komponenata fotografije, koje su korisne u predstavljanju i opisu oblika;
8. Segmentacija – vrši podelu fotografije na segmente ili objekte;
9. Predstavljanje i deskripcija – ova kategorija skoro uvek prati izlaz iz kategorije segmentacije, koji najčešće predstavlja neobrađeni podatak o pikselima

koji čini ili granicu segmenta ili sve tačke u samom segmentu. Deskripcija je proces izdvajanja atributa koji rezultiraju određenim kvantitativnim informacijama od interesa ili predstavljaju osnov za razlikovanje jedne klase objekata od druge;

10. Prepoznavanje objekata – proces koji dodeljuje oznaku objektu na osnovu njegovog deskriptora.

#### 4. ALGORITAM DETEKCIJE OBJEKATA NA DIGITALNIM FOTOGRAFIJAMA

Detekcija objekata je zadatak digitalne obrade fotografije i kompjuterske vizije koji otkriva (semantičke) instance objekata koje odgovaraju datom tipu (npr. lica, ljudi, vozila i zgrade) na fotografijama.

Klasifikacija, lokalizacija i detekcija objekata su, iako se odnose na različite zadatke, međusobno povezani algoritmi [6].

Klasifikacija objekata ima za cilj da predvidi oznaku klase objekta na fotografiji, dok se lokalizacija objekta bavi identifikovanjem lokacija i crtanjem graničnih okvira oko objekta na fotografiji [6].

Slika 2 predstavlja vizuelni prikaz klasifikacije objekta, dok se lokalizacija objekta prikazuje na slici 3.



CAT

SLIKA 2. PRIMER KLASIFIKACIJE OBJEKTA



SLIKA 3. PRIMER LOKALIZACIJE OBJEKTA

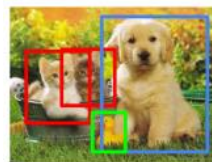
Detekcija objekata kombinuje ova dva algoritma, na način da lokalizaciju koristi za lociranje i iscertavanje graničnog okvira oko svakog objekta od interesa na fotografiji i svakom od njih dodeljuje oznaku klase, tako da koristi algoritam klasifikacije [6].

Na slici 4 dat je prikaz detekcije jednog objekta na fotografiji, a slika 5 prikazuje detekciju više objekata.



CAT

SLIKA 4. PRIMER DETEKCIJE JEDNOG OBJEKTA



CAT, DOG, DUCK

SLIKA 5. PRIMER DETEKCIJE VIŠE OBJEKATA

#### 4.1. UPOTREBA ALGORITMA DETEKCIJE OBJEKATA

Algoritam detekcije objekata koristi se u velikom broju industrija, a slučajevi upotrebe se kreću od lične bezbednosti do produktivnosti na radnom mestu. Otkrivanje i prepoznavanje objekata primenjuje se u mnogim oblastima računarske vizije, kao što su preuzimanje fotografija, bezbednost, nadzor, automatizovani sistemi vozila i pregled mašina.

#### 4.2. VRSTE ALGORITMA DETEKCIJE OBJEKATA

##### 4.2.1. YOLO

YOLO predstavlja skraćenicu engleskog izraza „*You only look once*” i jedan je od najpreciznijih i najtačnijih dostupnih algoritama za detekciju objekata. Napravljen je na osnovu izmenjene i prilagođene arhitekture Darknet. Prva verzija je inspirisana Google Net-om, koji je koristio tenzor za uzorkovanje fotografije i predvideo je sa najvećom tačnošću.

Tensor se generiše na osnovu slične procedure i strukture koja se takođe vidi u oblasti interesa koja je objedinjena i sastavljena da smanji broj pojedinačnih proračuna i učini analizu bržom.

Sledeća generacija je koristila arhitekturu sa samo 30 konvolucionih slojeva, koja se zauzvrat sastojala od 19 slojeva iz DarkNet-19 i dodatnih 11 za detekciju prirodnih objekata ili objekata u prirodnom kontekstu. Omogućio je precizniju detekciju i dobru brzinu, iako se borio sa fotografijama malih objekata i malih piksela. Verzija 3 je bila najtačnija verzija YOLO-a koja ima široku upotrebu zbog visoke stope preciznosti, kojoj je doprinela i arhitektura sa više slojeva [7].

##### 4.2.2. R-CNN

R-CNN predstavlja konvolucione neuronske mreže zasnovane na regionima. Ovaj model kombinuje predloge regiona za segmentaciju objekata i CNN-ove velikog kapaciteta za detekciju objekata [8]. Algoritam originalne R-CNN tehnike se odvija u 3 koraka:

1. Koristeći selektivni algoritam pretraživanja, nekoliko predloga regiona kandidata se izdvaja iz ulazne fotografije. U ovom algoritmu, u početnoj pod-segmentaciji generišu se brojni regioni kao kandidati. Zatim se slični regioni kombinuju da formiraju veće regione pomoću „pohlepno”



algoritma. Ovi regioni čine konačne predloge za region.

2. CNN komponenta iskrivljuje predloge i izdvađa različite karakteristike kao vektorski izlaz.
3. Osobine koje se izdvajaju unose se u SVM (engl. *Support Vector Machine*) za detekciju objekata od interesa u predlogu.

#### 4.2.3. FAST R-CNN

*Fast R-CNN* brži je od svog prethodnika jer nije potrebno unositi velik broj predloga kao ulaz u CNN po izvršenju. Operacija konvolucije se vrši da bi se generisala mapa karakteristika samo jednom po fotografiji. I *Fast R-CNN* i njegov prethodnik koristili su „*Selective Search*“ kao algoritam za određivanje predloga regiona [9].

Rad *Fast RCNN*-a, koji podrazumeva da se ulazna fotografija i više regiona od interesa (RoIs) unose u konvolucionu mrežu. Svaki RoI se udružuje u mapu karakteristika fiksne veličine, a zatim se potpuno povezanim slojevima (FC) preslikava u vektor obeležja. Mreža ima dva izlazna vektora po RoI-u: Softmax verovatnoće i regresioni pomaci ograničavajućeg okvira po klasi. Arhitektura je obučena po „*end-to-end*“ principu sa gubitkom u više zadataka.

#### 4.2.4. FASTER R-CNN

*Faster R-CNN* je još jedan model detekcije objekata zasnovan na regionu koji su Girshick i saradnici predložili kao poboljšanje R-CNN-a i *Fast R-CNN*. *Faster R-CNN* smanjuje vreme izvršavanja detekcije (npr. za sporiji R-CNN model) uvođenjem udruživanja povraćaja ulaganja, ali ipak izračunavanje predloga regiona postaje usko grlo. *Faster R-CNN* uvodi mrežu regionalnih predloga (RPN, engl. *Region Proposal Network*). Postiže gotovo besplatne predloge za region deleći konvolutivne karakteristike sa mrežom za otkrivanje [8].

Mreža za predloge regiona (RPN) je FCN (engl. *Fully Convolutional Network*) koja predviđa regije koje potencijalno sadrže objekat sa okvirima koji ograničavaju objekte zajedno sa ocenama objektivnosti (odnosno verovatnoćom da region sadrži objekat) na svakoj poziciji. Obuka RPN-a po „*end-to-end*“ principu omogućava joj da visokokvalitetno predvidi predloge za region. Tada *Faster R-CNN* koristi ove regione za moguće otkrivanje [9]. RPN i *Fast R-CNN* spojeni su u jednu mrežu. Mreža je zajednički obučena sa četiri gubitka:

1. RPN pruža klasifikaciju objekta/ne-objekta (sa ocenom objektivnosti).
2. RPN koristi regresiju za izračunavanje koordinata polja.
3. Konačni klasifikator (iz *Fast R-CNN*) klasifikuje objekat (sa ocenom klasifikacije).
4. Izlazni ograničavajući okviri koji odgovaraju objektu izračunavaju se regresijom.

## 5. ZAKLJUČAK

U radu je obrađena tema algoritama detekcije objekata pri digitalnoj obradi fotografije. Prikazani su osnovni koncepti računarske vizije i digitalne obrade fotografija, kao i njihova osnovna namena. Objašnjena je osnovna organizacija koraka pri digitalnoj obradi fotografije, koja obuhvata 10 koraka, čije izvršenje zavisi od namene digitalne obrade fotografije. Opisana je glavna razlika između algoritama klasifikacije, lokalizacije i detekcije jednog ili više objekata na fotografijama, koja podrazumeva značenje algoritma detekcije. Nakon toga, predstavljene su neke od postojećih vrsta prethodno navedenog algoritma, koje imaju određene razlike u samoj arhitekturi modela, što dovodi do drugačijih rezultata uz parametre tačnosti, brzine i performansi.

## 6. LITERATURA

- [1] Demush, R. (2019, February 26). A Brief History of Computer Vision (and Convolutional Neural Networks). Hacker Noon.
- [2] Ptucha, R., Petroski Such, F., Pillai, S., Brockler, F., Singh, V., & Hutkowski, P. (2019). Intelligent character recognition using fully convolutional neural networks. *Pattern Recognition*, 88, 604–613.
- [3] Prince, S. J. D. (2012). *Computer Vision: Models, Learning, and Inference* (1st ed.). Cambridge University Press.
- [4] Jensen, J R. (1986) *Introductory digital image processing: A remote sensing perspective*. United States.
- [5] Ribeiro, Sergio. (2014). USING SIMPLCV FOR SEED METADATA EXTRACTION INTO XML DOCUMENT. *Iberoamerican Journal of Applied Computing*, 4. 29.
- [6] Dey, S., 2020. *Python Image Processing Cookbook*. Birmingham: Packt Publishing, Limited.
- [7] H C, D. (2020, June). An Overview of You Only Look Once: Unified, Real-Time Object Detection. *International Journal for Research in Applied Science and Engineering Technology*, 607–609.
- [8] Sundaramoorthy, K., G, M., & Marimuthu, R. (2019, July). OPTIMAL AND FAST HAND GESTURE RECOGNITION MODEL USING FASTER R-CNN. *Recent Patents on Computer Science*.
- [9] Object Detection and Tracking using Faster R-CNN. (2019, September). *International Journal of Recent Technology and Engineering*, 4894–4900.

### Kratka biografija:



**Katarina Gavrilov** rođena je u Novom Sadu 1996. godine. Master rad je odbranila 2021. godine na Fakultetu tehničkih nauka iz oblasti inženjerstva informacionih sistema.