Zbornik radova Fakulteta tehničkih nauka, Novi Sad

UDK: 004.9
DOI: https://doi.org/10.24867/13BE41Pantovic

DETEKCIJA OBJEKATA U SAOBRAĆAJU UPOREBOM KONVOLUCIONIH NEURONSKIH MREŽA

OBJECT DETECTION IN TRAFFIC SCENES WITH CONVOLUTIONAL NEURAL NETWORKS

Sofija Pantović, *Fakultet tehničkih nauka, Novi Sad*

Oblast – ELEKTROTEHNIKA I RAČUNARSTVO

Kratak sadržaj – *Detekcija objekata je ključna tehnologija koja stoji iza naprednih sistema za asistenciju tokom vožnje. U ovom radu je prikazana primena modela konvolucionih neuronskih mreža u rešavanju problema detekcije objekata od interesa iz sekvence slika. Obeležene slike koje sadrže različite scenarije zabeležene tokom dnevne gradske vožnje preuzete su iz CrowdAI baze [1]. Slike su korišćene za učenje modela i ocenu njegovih performansi tokom faze testiranja. Analizirani su rezultati dobijeni korišćenjem detektora sa različitim brojem konvolucionih slojeva i različitim aktivacionim funkcijama neurona u cilju primene ovakvih modela za detekciju objekata učesnika u saobraćaju u realnom vremenu.*

Ključne reči: *Detekcija objekata, Klasifikacija, Konvoluciona neuronska mreža, Okvir*

Abstract – *Object detection is the key technology behind advanced driver assistance systems. This paper demonstrates an application of convolutional neural networks in the task of detecting objects of interest from the sequence of images. Labeled images with various driving scenarios recorded during daylight city drive are taken from CrowdAI database [1]. Images were used for model training and evaluation during testing phase. Results obtained using models with different number of convolutional layers and various activation functions are analyzed with purpose of using these models for real-time object detection.*

Keywords: *Bounding box, Classification, Convolutional neural network, Object detection*

1. UVOD

Detekcija objekata iz sekvence slika je ključna tehnologija koja stoji iza naprednih sistema za asistenciju tokom vožnje (engl. *advanced driver assistance systems*). Sistemi za asistenciju tokom vožnje detektuju vozne trake, ivice puta, druga vozila ili pešake, sa ciljem poboljšanja sigurnosti putnika i drugih učesnika u saobraćaju.

Klasifikacija objekata je proces predviđanja klase objekta sa slike. Lokalizacija objekata odnosi se na identifikaciju lokacije jednog ili više objekata u slici i određivanje njihovog okvira (engl. *bounding box*).

NAPOMENA:
Ovaj rad proistekao je iz master rada čiji mentor je dr Vladimir Bugarski, docent.

Kombinacijom ovih zadataka nastaje prepoznavanje tj. detekcija objekata iz slike [2]. Konvolucione neuronske mreže (engl. *convolutional neural networks*) su se pokazale kao uspešne kod prepoznavanja objekata u slici, pa se nameću kao logičan izbor za rešenje problema detekcije [3].

CrowdAI baza podataka od preko 65.000 obeležja iz 9.423 kadra visoke rezolucije korišćena je za obuku neuronske mreže. Skup podataka je obeležen kombinacijom ručnog unosa i tehnika mašinskog učenja. Objekti od interesa pripadaju sledećim klasama:

1. Automobil;
2. Kamion;
3. Pešak.

Pored oznake klase, obeležje sadrži koordinate okvira detektovanog objekta u formatu $[x_{min}, y_{min}, x_{max}, y_{max}]$.

2. KONVOLUCIONA NEURONSKA MREŽA

Konvoluciona neuronska mreža je algoritam dubokog učenja koji kao ulaz, između ostalog, prima sliku, dodeljuje značaj (naučene težine i pristrasnost (engl. *bias*)) različitim objektima u slici, kako bi bio u mogućnosti da ih razlikuje [4]. Konvolucione neuronske mreže se, kao i regularne višeslojne neuronske mreže, sastoje od jednog ulaznog, jednog izlaznog i barem jednog ili više skrivenih slojeva. Ono što razlikuje konvolucionu neuronsku mrežu od obične jesu konvolucioni slojevi i slojevi sažimanja.

2.1 Konvolucioni sloj

Konvolucija dve funkcije $f, g: \mathbb{R}^d \rightarrow \mathbb{R}$ definisana je kao:

$$(f * g)(\boldsymbol{x}) = \int f(\boldsymbol{z}) g(\boldsymbol{x} - \boldsymbol{z}) d\boldsymbol{z} \qquad (1)$$

U slučaju diskretnog skupa i dvodimenzionih tenzora, integral se pretvara u sumu:

$$(f * g)(i, j) = \sum_a \sum_b f(a, b) g(i - a, j - b) \qquad (2)$$

gde a i b označavaju granice konvolucionog kernela u odnosu na centar.

Tokom faze propagiranja signala unapred, ulazna slika se konvoluira sa kernelom tj. filterom, što jasno ukazuje da je zadatak obuke konvolucione mreže podešavanje težina u okviru različitih kernela. Proces konvolucije omogućava redukciju ulaznih slika u formu koja je lakša za obradu, bez gubitaka obeležja koja su esencijalna za postizanje dobre predikcije [4]. Uprošćena reprezentacija ulaznih slika se naziva aktivaciona mapa ili mapa obeležja.

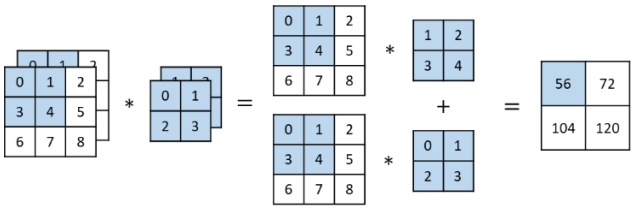
Za slučaj višestrukih kanala ulaza $\boldsymbol{X}$ i skrivene mape obeležja $\boldsymbol{H}$ važi relacija:

$$[\boldsymbol{H}]_{i,j,d} = \sum_{a=-\Delta}^{\Delta} \sum_{b=-\Delta}^{\Delta} [\boldsymbol{V}]_{a,b,c,d} [\boldsymbol{X}]_{i+a,j+b,c} \qquad (3)$$

gde je $[\boldsymbol{V}]_{a,b,c,d}$ tenzor težinskih faktora u okviru kernela, $[\boldsymbol{X}]_{i+a,j+b,c}$ deo ulazne slike obuhvaćene kernelom, dok d ukazuje na broj izlaznih kanala skrivenih mapa obeležja [4].

2.2 Primena konvolucije na slikama

Kada ulazne slike sadrže višestruke kanale potrebno je konstruisati konvolucioni kernel sa istim brojem ulaznih kanala. Skrivena mapa obeležja se formira sabiranjem rezultata konvolucije dvodimenzionih tenzora ulaza i tenzora kernela za svaki od kanala (slika 1).



Slika 1. *Konvolucija višestrukih slojeva – primer sa dva ulazna kanala i jednim izlaznim*

2.3 Sloj sažimanja

Tokom propagacije signala unapred kroz mrežu, prostorna rezolucija se postepeno smanjuje, informacije se združuju tako da što je signal dublje u mreži, veće je receptivno polje ulaza na koje je skriveni sloj osetljiv [5]. Operacija sažimanja se primenjuje identično kao i konvolucioni kernel, sa razlikom što su operatori sažimanja deterministički tj. ne sadrže parametre. Najčešće se implementiraju tako da računaju maksimalnu ili prosečnu vrednost elemenata zahvaćenih prozorom. Sloj sažimanja obrađuje ulazne kanale pojedinačno, te će za svaki ulazni kanal ovaj sloj na svom izlazu izbaciti njemu odgovarajuću izlaznu aktivacionu mapu.

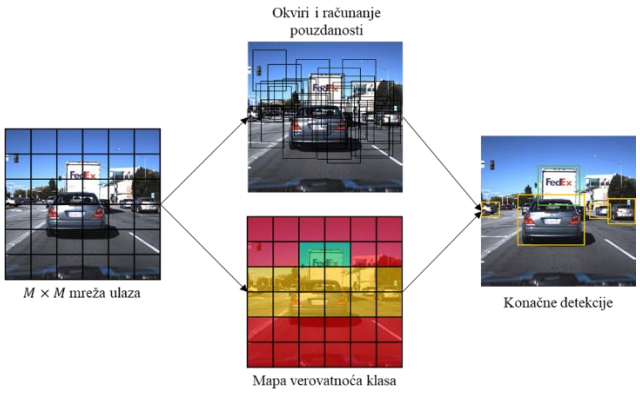
2.4 Arhitektura mreže

Tokom jednog prolaska kroz mrežu, model YOLO [4] arhitekture prikuplja značajna obeležja iz ulaznih slika i kasnije ih koristi za estimaciju okvira i klasa objekata u njima. YOLO model paralelno, za veći broj objekata u slici, estimira višestruke okvire i klasne verovatnoće za svakog od njih [6].

U poređenju sa sistemima slične namene, YOLO arhitektura, zbog jednostavnosti konfiguracije (ulazna slika kroz mrežu prolazi samo jednom), ima dobre performanse u realnom vremenu, čak i bez korišćenja naprednijih grafičkih komponenti [6].

Princip rada YOLO modela prikazan je na slici 2. Ulazna slika deli se na MxM jednakih ćelija. Unapred određen broj početnih okvira N koristi se za estimaciju objekata u svakoj od ćelija.

Pouzdanosti estimacije objekata u ćelijama formiraju mapu verovatnoća, koja govori da li se unutar okvira nalazi objekat ili ne, tj. koliko je model zapravo siguran u određenu predikciju. Na kraju se iz mapa verovatnoća klasa izdvajaju konačne detekcije objekata.



Slika 2. Princip rada YOLO modela.

2.5 Definicija okvira i računanje pouzdanosti

Jedan okvir definiše pet parametara: t_x, t_y, t_w, t_h i rezultat pouzdanosti. Koordinate piksela koji određuje gornji levi ugao okvira su date uređenim parom parametara (t_x, t_y). Parametri t_w i t_h diktiraju visinu i širinu okvira, respektivno. Peti parametar, rezultat pouzdanosti, sadrži informacije o verovatnoćama za svaku od klasa u modelu. Zadatak konvolucione mreže za detekciju objekata jeste estimacija ovih pet parametara. YOLO estimira tenzor veličine $M * M * (N \cdot 5 + C)$ za svaku sliku skupa za obuku, gde je $M * M$ broj ćelija, $N \cdot 5$ broj apriornih okvira po ćeliji sa pet parametara za svaki okvir, dok je C broj klasa.

YOLO model vrši predikciju objekata u završnim slojevima mreže na osnovu klasifikacionih i lokalizacionih grešaka okvira, tj. grešaka između tačne i estimirane pozicije. Rezultat pouzdanosti p određene klase C jednak je umnošku verovatnoće postojanja objekta te klase unutar okvira i odnosa preseka i unije (engl. *intersection over union* - IOU) između apriornog i stvarnog okvira:

$$p(C) = p(objekta) \cdot IOU \qquad (4)$$

Detektovani okvir koji se savršeno poklapa sa originalnim imaće izlaz 1, dok će svako odstupanje u preklapanju proizvesti manji IOU rezultat.

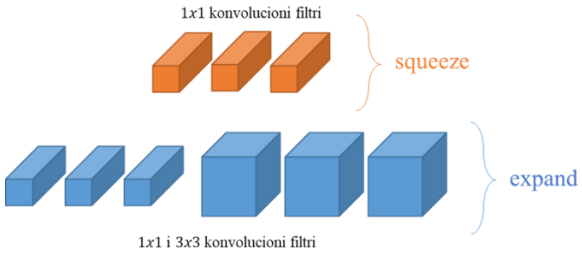
3. SQUEEZENET MODEL

Arhitektura neuronske mreže SqueezeNet koja čini unapređenu verziju YOLO modela (YOLOv3) [6], izabrana je za obuku nad CrowdAI bazom podataka. SqueezeNet mrežu karakteriše arhitektura sa minimalnim brojem parametra uz održavanje zadovoljavajućeg nivoa tačnosti.

U [6] i [7] su navedene strategije koje su korišćene kako bi se postiglo smanjenje parametara modela:

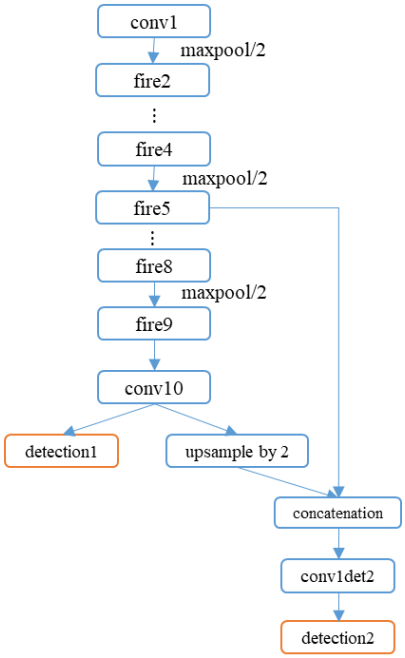
1. Umesto filtara veličine $3x3$, u konvolucionim slojevima su korišćeni filtri veličine $1x1$, čime se broj parametara smanjuje devet puta.
2. Upotrebom slojeva za sažimanje (engl. *squeeze*) je redukovan broj ulaznih kanala u filtre veličine $3x3$, što značajno smanjuje broj potrebnih parametara.
3. Smanjenje dimenzionalnosti u kasnijim nivoima mreže realizovano je postavljanjem koraka na vrednost veću od 1 u konvolucionim i slojevima sažimanja.

Predložene strategije smanjenja parametara su realizovane formiranjem okidačkog modula (engl. *fire modul*) (slika 3). Ovaj modul se sastoji od konvolucionog sloja sa $1x1$ filtrima koji predstavlja sažimajući sloj, čiji se izlazi prosleđuju u sloj koji je kombinacija $1x1$ i $3x3$ konvolucionih filtara, takozvani proširujući sloj (engl. *expand*).



Slika 3. *Organizacija konvolucionih filtara u okidačkom modulu.*

Model SqueezeNet arhitekture prikazan je na slici 4.



Slika 4. *SqueezeNet arhitektura sa 2 detekcione glave.*

Počinje jednostavnim konvolucionim slojem nakon kog sledi 8 okidačkih modula, kod kojih se broj filtara postepeno povećava od početka ka kraju mreže. Sažimanje maksimumom realizovano je nakon početnog konvolucionog sloja, fire4 i fire8 modula.

Na završni konvolucioni sloj modela povezan je novi sloj proširen faktorom 2, koji se ujedno spaja i sa prethodnom mapom obeležja istih dimenzija. Na ovaj način formirane su dve tzv. detekcione glave, od kojih je jedna dvostruko manja i bolje detektuje sitnije objekte u slici. Broj i dimenzije detekcionih glava u korelaciji su sa raznolikošću i dimenzijama objekata koji se detektuju.

3.1. Generisanje okvira i priprema baze podataka

Dimenzije apriornih okvira određuju se algoritmom grupisanja *k*-najbližih suseda nad okvirima objekata iz skupa za obuku [5]. Analiziraju se svi okviri trening skupa, a kao rezultat generiše se k odabranih okvira koji najbolje odgovaraju njihovim dimenzijama.

Kako je cilj odrediti dimenzije početnih okvira koji vode do dobrog rezultata preklapanja, za računanje udaljenosti koristi se mera koja ne zavisi od veličine okvira (5).

$$d(okvir, centroid) = 1 - IOU(okvir, centroid) \qquad (5)$$

Model detektora implementiran je korišćenjem programske platforme MATLAB. Ova platforma nudi različite modele konvolucionih neuronskih mreža sa već podešenim vrednostima težinskih faktora, naučenih tokom obuke nad Imagenet bazom podataka. Unapred naučene težine pomažu procesu obuke sopstvene mreže tako što ubrzavaju konvergenciju, kao i učenje. Nakon podele uzoraka na skupove, implementirana je funkcija za proveru validnosti uzoraka. Uzorci sa neodgovarajućim formatom slike, neočekivanim vrednostima koordinata okvira ili pogrešnim formatom labela isključeni su iz razmatranja. Proširivanje raznolikosti test uzoraka (engl. *data augmentation*) doprinosi povećanju tačnosti mreže. Implementirane su varijacije boja, horizontalne rotacije i skaliranje za 10%.

4. REZULTATI

CrowdAI baza sadrži 9.423 slike. Skup podataka podeljen je u odnosu 70:30 na trening i test skup, respektivno, te je u obuci učestvovalo 6.657 slika, dok je za procenu tačnosti modela korišćeno 2.766 uzoraka. Kako uzorci test skupa nisu učestvovali u obuci, nije postojala bojazan da su parametri mreže prilagođeni uzorcima koji se korišćeni za evaluaciju performansi.

Originalni uzorci test skupa su slike veoma visoke rezolucije (1920x1200x3). Početni pokušaji obuke konvolucionih mreža, koje prihvataju ulazne slike datih dimenzija, bili su neuspešni. Arhitektura, odabrana teorijskom analizom, pokazala se kao neadekvatna. Tokom procesa obuke, već u prvoj epohi, uočeno je da greška mreže ima nagli porast, čime je dalji trening mreže postao beskoristan. Uz pretpostavku da je arhitektura mreže previše jednostavna, pokušalo se sa povećanjem broja konvolucionih slojeva i povećanjem broja filtara u okviru njih. Međutim, pretpostavka da jednostavnost mreže dovodi do loših rezultata nije ni potvrđena, niti opovrgnuta. Naime, memorijski i procesorski kapaciteti računara korišćenog za istraživanje nisu bili dovoljni da se obuka završi. Stoga, ulazni frejmovi visoke rezolucije su bili komprimovani u format (227x227x3).

4.1 Evaluacione mere

Glavne evaluacione mere kod detekcije objekata su preciznost i osetljivost. Preciznost modela govori koliki je udeo tačno klasifikovanih objekata u skupu klasifikovanih, dok osetljivost govori koliki je udeo tačno klasifikovanih uzoraka u skupu svih pozitivnih uzoraka. Mera koja povezuje preciznost i osetljivost naziva se prosečna preciznost (engl. *average precision*).

Prosečna preciznost predstavlja srednju vrednost preciznosti dobijenih za N jednako razmaknutih pragova za pouzdanost. U izrazu (6), AP predstavlja prosečnu pouzdanost, N je broj pragova, a P obeležava preciznost.

$$AP = \frac{1}{N}\sum_{k=0}^{N-1} P\left[r = \frac{k}{N-1}\right] \qquad (6)$$

Mera koja opisuje kvalitet nekog detektora je srednja prosečna preciznost mAP (engl. *mean average precision*). Srednja prosečna preciznost predstavlja srednju vrednost prosečnih preciznosti za sve klase koje model detektuje nad celokupnim test skupom (7).

$$mAP = \frac{1}{Q}\sum_{i=0}^{Q-1} AP_i \qquad (7)$$

4.2 Poređenje rezultata mreža sa različitim brojem apriori okvira

U ovom eksperimentu korišćena je opisana arhitektura SqueezeNet konvolucione mreže sa dve detekcione glave. Mreža se sastoji od 13 konvolucionih slojeva, ne računajući aktivacione mape u detekcionim glavama. Unutar svih slojeva mreže korišćene su tri različite aktivacione funkcije: ReLU, leaky ReLU i hiperbolički tangens. U sloju sažimanja korišćena je metoda sažimanja maksimalnom vrednošću sa korakom 2.

Rezultati modela postignuti tokom 30 epoha za slučaj 11 i 5 apriornih okvira predstavljeni su u Tabeli 1 preko vrednosti prosečnih preciznosti po klasama i srednje prosečne preciznosti modela.

Tabela 1. *Rezultati detektora sa različitim brojem apriori okvira.*

Broj okvira	AP			mAP
	Automobil	Kamion	Pešak	
11	84,31%	57,35%	88,27%	76,64%
5	80,50%	41,09%	89,58%	70,39%

Naredna zamisao u pogledu poboljšanja tačnosti detektora je promena korišćenih aktivacionih funkcija. Ovde su pored ReLU aktivacione funkcije, performanse modela sa 11 apriori okvira evaluirane za hiperbolički tangens (HT) i leaky ReLU aktivacione funkcije.

Tabela 2. *Rezultati detektora sa različitim aktivacionim funkcijama.*

	AP			mAP
	Automobil	Kamion	Pešak	
ReLU	84,31%	57,35%	88,27%	76,64%
Leaky ReLU	85,07%	61,04%	95,99%	80,70%
HT	79,46%	19,21%	93,11%	63,93%

5. ZAKLJUČAK

Detektor objekata se u slučaju sa 11 apriori okvira, prema srednjoj prosečnoj preciznosti, pokazao kao uspešniji u odnosu na model sa 5 apriori okvira za oko 6%. Iz tog razloga je dalje istraživanje nastavljeno sa prvo pomenutim modelom. Leaky ReLU aktivaciona funkcija najbolje se pokazala u slučaju sve tri klase. Pokazuje dobre performanse čak i u slučaju klasa sa manjim brojem predstavnika u skupu podataka (Kamion, Pešak). Neki od pravaca daljeg istraživanja su testiranje modela na većim skupovima za obuku i sa kompleksnijim arhitekturama mreže.

6. LITERATURA

[1] L. Hunjung, „Awesome vehicle datasets“ (CrowdAI baza podataka), https://github.com/hunjung-lim/awesome-vehicle-datasets

[2] J. Brownlee, „A Gentle Introduction to Object Recognition With Deep Learning“, https://machinelearningmastery.com/object-recognition-with-deep-learning. (poslednji pristup u martu 2021. godine)

[3] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, “Gradient-based learning applied to document recognition”, *Proceedings of the IEEE*, 86(11), pp. 2278–2324, 1998.

[4] A. Zhang, Z.C. Lipton, M. Li, A.J. Sola, „Dive into Deep Learning“, https://d2l.ai/ (poslednji pristup u martu 2021. godine)

[5] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, “You Only Look Once: Unified, Real-Time Object Detection”, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[6] F.N. Iandola, S. Han, M.W. Moskewicz, K. Ashraf, W.J. Dally, K. Keutzer, “SqueezeNet: AlexNet-Level Accuracy with 50x Fewer Parameters and <0.5MB Model Size”, arXiv Prepr. arXiv1602.07360, 2016.

[7] K. He, J. Sun, “Convolutional neural networks at constrained time cost“, *Conference on Computer Vision and Pattern Recognition(CVPR)*, 2015.

Kratka biografija:

Sofija Pantović rođena je u Vrbasu 1995. god. Master rad na Fakultetu tehničkih nauka iz oblasti Elektrotehnike i računarstva – Automatika i upravljanje sistemima odbranla je 2021.god.

kontakt: sofija.pantovic33@gmail.com