

Automatska detekcija oštećenja na putu pomoću dubokih neuralnih mreža

David Đukić, Goran Ferenc, Stefan Stefanović

SYRMIA doo Novi Sad

Beograd, Srbija

david.djukic@Syrmia.com, goran.ferenc@Syrmia.com, stefan.stefanovic@Syrmia.com

Sažetak— Putna infrastruktura je ključno javno dobro koje doprinosi ekonomskom razvoju i rastu svake države i istovremeno pruža razne društvene pogodnosti. Međutim, površina puta se oštećuje i propada tokom vremena zbog raznih faktora kao što su starost, obim saobraćaja, vremenski uslovi, inženjerski pristup i materijali koji su korišćeni tokom njegove izgradnje. Stoga znanje o stepenu propadanja puta pomoću automatske detekcije rupa je ključno za efikasno i isplativo održavanje, gde je cilj očuvanje dobrog i pre svega bezbednog stanja na putu. Detekcija objekata je kompleksan problem, jer zahteva da model može da detektuje više objekata na slici, koji mogu biti različite veličine, dok je istovremeno potrebno izvršiti lokalizaciju i klasifikaciju objekata. Takođe, izazov predstavlja postići balansiranost između tačnosti modela i brzine izvršavanja. Opisana je familija metoda koje koriste konvolucione neuralne mreže sa predlozima regija i implementiran je Brži R-CNN nad bazom slika koje sadrže nekoliko hiljada rupa. Analizirani su rezultati na test skupu podataka, gde se model pokazao kao dosta dobar u detekciji rupa na putu i dobijena vrednost F1 skora je 0.697.

Ključne reči— Duboko učenje; Detekcija objekata; Konvolucione neuralne mreže; Brži R-CNN;

I. UVOD

Prema studiji koju su sproveli Miller i Zalosnja [1], u Sjedinjenim Američkim Državama loše stanje puta uzrokuje 52.7% smrtnih slučajeva na godišnjem nivou (koji nastaju kao posledica nesreće na putu) i 38% nesreća bez smrtnog ishoda. Ovako izazvane nesreće koštaju američku ekonomiju više od 217 milijardi dolara svake godine, što je 3.5 puta veći iznos novca od godišnjeg ulaganja za kapitalna poboljšanja na putevima. Procenjuje se da američka preduzeća imaju oko 22 milijarde dolara godišnjih ekonomskih troškova zbog nesreća na putu u kojima su učestvovali njihovi zaposleni, pri čemu je razlog nastanka nesreće – loše stanje na putu.

Oštećenja na putu se obično detektuju korišćenjem jednog od tri pristupa: ručni, poluautomatski ili potpuno automatizovani. Kod ručnog pristupa, stručno lice obavlja vizuelnu inspekciju hodanjem po površini puta ili posmatranjem iz vozila koje se sporo kreće. Vizuelni pregled nije dobra praksa, jer zavisi od subjektivne procene inspektora i zahteva značajnu ljudsku intervenciju za koju se ispostavlja da oduzima mnogo vremena s obzirom na veliku dužinu putne mreže. Štaviše, inspektori moraju često biti fizički prisutni u saobraćajnoj traci, izlažući se opasnim uslovima. Kod poluautomatskog pristupa, prikupljaju se slike iz vozila koje se brzo kreće, ali identifikacija

oštećenja se radi na radnim stanicama u kancelariji. Ovaj pristup poboljšava bezbednost, ali i dalje koristi ručnu identifikaciju, što oduzima mnogo vremena [2].

Veliki je broj radova koji su se bavili detekcijom oštećenja na putu, u kojima se primenjuju različite tehnike: detekcija pomoću senzora, 3D rekonstrukcija (bazirano na stereoviziji), detekcija pomoću digitalne obrade slike i tehnike zasnovane na veštačkoj inteligenciji (mašinsko učenje i duboke neuralne mreže). Sistemi koji se oslanjaju na automatsku detekciju pomoću senzora imaju ograničenja poput: senzori se mogu oštetiti zbog uslova na putu, ne postoje informacije o površini i obliku oštećenja, za detekciju je neophodno preći tačno preko oštećenja, što znači da će postojati i veliki broj nedetektovanih rupa. Stereovizijske tehnike nisu efikasne, jer su veliki komputacioni zahtevi prilikom rekonstrukcije puta i potrebno je repetitivno poravnavanje kamera da bi se postigli dobri rezultati. Detekcija korišćenjem tehnika iz digitalne obrade slike nije robustna, jer zahteva prilagođavanje parametara različitim uslovima na putu i često nije uopšte primenljiva u realnom vremenu [3].

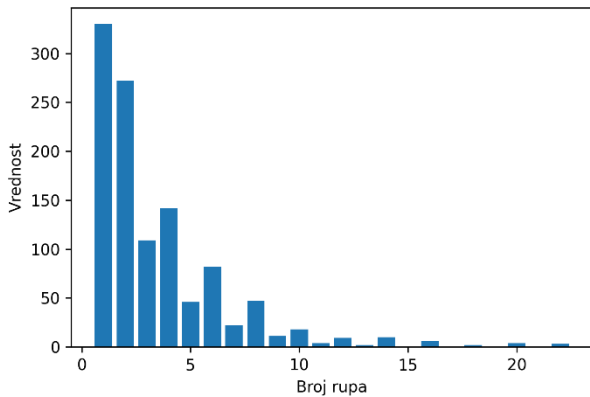
Potpuno automatizovan pristup često koriste vozila opremljena sofisticiranim i skupim sensorima kao što su laseri, profilari puta, lidar itd. Takva vozila su skupa, a cena kupovine tih sistema može dostići pola miliona dolara u zavisnosti od senzora. Kako su tokom poslednje decenije razvijene digitalne kamere sa visokom rezolucijom, danas je moguće postići efikasnu i isplativu inspekciju puteva samo pomoću kamere u automobilu. Od svih tehnika za automatsku detekciju, po pitanju tačnosti i brzine izvršavanja izdvajaju se metode koje su bazirane na tehnikama mašinskog učenja i dubokog učenja. Razvoj dubokog učenja je uticao na skoro svaki aspekt kompjuterske vizije: klasifikacija objekata, segmentacija slika, simultana lokalizacija i mapiranje samo su neke oblasti kompjuterske vizije koje su se značajno promenile sa razvojem dubokog učenja.

Cilj ovog rada je automatska detekcija oštećenja na putu korišćenjem dubokog učenja i Bržeg R-CNN modela na osnovu snimaka kamere iz automobila. U nastavku je detaljno opisana baza podataka, kao i teorijski pregled modela za detekciju objekata, gde je poseban akcenat stavljen na dvostepene detektore koji se baziraju na konvolucionim neuralnim mrežama sa predlozima regija. U četvrtom poglavlju je opisana konkretna implementacija, dok su u petom poglavlju prikazani rezultati

obučavanja, zatim korišćene metrike za evaluaciju modela i dobijeni rezultati na test skupu podataka.

II. BAZA PODATAKA

U ovom radu korišćena je javno dostupna baza podataka. Celokupni podaci su podeljeni na trening, validacioni i test skup u razmeri 70:15:15. Trening podaci sadrže 1119 slika sa rupama na putu, pri čemu se na svakoj slici nalazi različiti broj rupa (od 1 do 22), što je prikazano na Sl. 1. Za svaku sliku iz trening, validacionog i test seta dostupni su fajlovi u kojima je data pozicija gornjeg levog ugla, visina i širina graničnog okvira za svaku rupu na putu.



Slika 1. Histogram broja rupa na trening setu podataka

Primer slike iz trening seta sa označenim graničnim okvirom (crveni pravougaonici) oko svakog objekta od interesa je prikazan na Sl. 2.



Slika 2. Primer slike iz trening seta sa graničnim okvirima objekata

III. DETEKCIJA OBJEKATA POMOĆU DUBOKIH NEURALNIH MREŽA

Detekcija objekata je tehnika kompjuterske vizije čiji je zadatak identifikacija i labeliranje objekata na slikama, videima, pa čak i na uživo snimcima. Prvi izazov kod detekcije objekata je promenljivi broj objekata. Prilikom obučavanja modela zasnovanih na mašinskom učenju, izlazi obično moraju da se prikažu kao vektori fiksne veličine. Pošto broj objekata na slici nije unapred poznat, ne može se odrediti tačan broj izlaza. Drugi izazov je različita veličina objekata, jer u nekim slučajevima objekti pokrivaju veći deo slike, ali se može desiti da neki objekti mogu biti značajno manji i jedva vidljivi na samoj slici.

Na kraju, možda i najveći izazov predstavlja istovremeno rešavanje dva zahteva - izvršiti lokalizaciju i klasifikaciju objekta.

Poslednju deceniju sa povećanjem komputacione moći računara i povećanjem broja labeliranih podataka, duboko učenje uspeva da postigne dosta veliku tačnost u različitim problemima koji zahtevaju detekciju objekata. Postoje dve kategorije modela za detekciju objekata baziranih na dubokim neuralnim mrežama: dvostepeni detektori i jednostepeni detektori. Dvostepeni detektori, u svom prvom delu, generišu potencijalne regione gde se objekti mogu nalaziti, dok u drugom delu određuju kojoj klasi svaki objekat pripada i granični okvir u kom se nalazi objekat. Kod dvostepenih detektora se izdvaja familija metoda koje koriste konvolucione neuralne mreže sa predlozima regija (engl. Region Based Convolutional Network – R-CNN), gde se posebno izdvaja Brzi R-CNN model. Jednostepeni detektori kao što su YOLO (engl. You Only Look Once) i SSD (engl. Single Shot Multibox Detector) uspevaju da predvide granični okvir objekta i verovatnoću da pripada nekoj klasi, bez da prethodno traže potencijalne regije gde se objekti mogu nalaziti. Razlika između ove dve kategorije detektora je to što dvostepeni detektori postižu veću tačnost prilikom detekcije, ali su sporiji i ne mogu se koristiti u realnom vremenu, dok su jednostepeni detektori značajno brži, ali uglavnom postižu nešto manju tačnost [3].

A. R-CNN i Brzi-R-CNN

Prvi rad u kojem je detaljno prikazan R-CNN model je publikovao Girshick 2014. godine [4]. Arhitektura R-CNN-a se sastoji iz četiri koraka:

1. Postavljanje slike na ulaz.
2. Ekstrakcija predloga regija pomoću Selektivne pretrage (generiše se oko 2000 predloga regiona).
3. Korišćenje transfera znanja i unapred obučene konvolucione neuralne mreže za dobijanje mape osobina svakog regiona.
4. Klasifikacija svakog regiona metodom potpornih vektora.

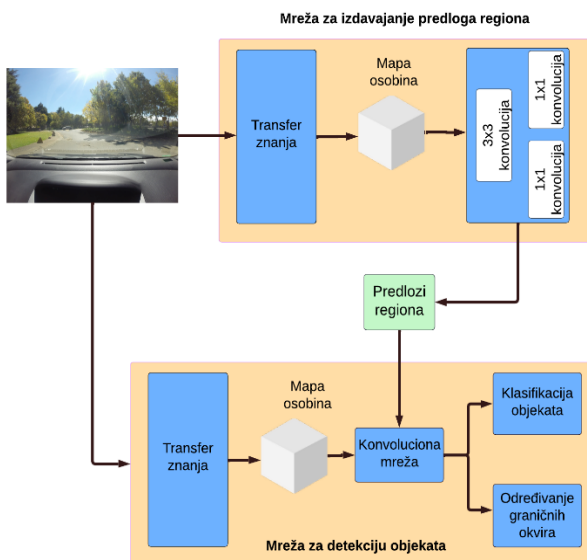
U prvom koraku se na ulaz algoritma dovodi slika. Zatim se izdvajaju predlozi regija pomoću algoritma Selektivne pretrage. Selektivna pretraga generiše podsegmentacije slike koje bi mogle da pripadaju jednom objektu na osnovu boje, teksture, veličine i oblika i iterativno kombinuje slične regione da bi se formirali novi objekti. Ovim se dobijaju predlozi regija različitih razmera, gde se generiše za svaku sliku oko dve hiljade predloga označenih pravougaonim graničnim okvirima. Svaki od dve hiljade predloga regiona koji je generisan iz slike se konvertuje jednostavnim reskaliranjem tako da bude fiksne veličine, kako bi se pripremio za ulaz na unapred istreniranu konvolucionu neuralnu mrežu (primenjuje se Transfer znanja). Izlaz iz mreže je 4096-dimenzioni vektor koji predstavlja mapu osobina za svaki predlog regiona. U ovom slučaju konvoluciona neuralna mreža služi kao ekstaktor osobina za svaki region. Zatim se na vektor osobina primenjuje metoda potpornih vektora da bi se klasifikovalo prisustvo objekta unutar tog predloga regiona. Pored predviđanja prisustva objekta, predviđaju se i četiri

vrednosti koje predstavljaju vrednosti pomeranja da bi se povećala preciznost graničnog okvira.

Iako je R-CNN značajno doprineo poboljšanju detekcije objekata, postoji više razloga zašto se ne koristi u praksi. Potrebna je velika količina vremena da se obuču neuralna mreža i ne može se implementirati u realnom vremenu, jer je potrebno oko 47 sekundi za svaku test sliku. Takođe, selektivna pretraga u nekim slučajevima može da generiše loše predloge regiona. Kao poboljšanje R-CNN-a, razvijen je Brzi R-CNN (engl. Fast R-CNN) [5]. Pristup je sličan kao kod R-CNN-a, s tim što se umesto predloga regiona, sada ulazna slika prvo propušta kroz unapred istreniranu mrežu, pa se na dobijenoj mapi osobina izdvajaju regioni. Razlog zašto je Brzi R-CNN brži od R-CNN-a je što ne mora da se za svaku sliku prosleđuje dve hiljade predloga regiona konvolucionoj neuralnoj mreži. Procenjuje se da je Brzi R-CNN 45 puta brži u testiranju i 9 puta brži u treniranju od R-CNN.

B. Brži R-CNN

Algoritmi poput R-CNN-a i Brzog R-CNN-a koriste selektivnu pretragu kako bi pronašli potencijalne regije gde se nalazi objekat. Selektivna pretraga je dugotrajan proces i utiče na performanse mreže (traje oko 2ms po slici). Iz tog razloga, osmišljen je novi model – Brži R-CNN (engl. Faster R-CNN) koji eliminiše selektivnu pretragu i omogućava mreži da nauči gde se nalaze regije od interesa [6]. Arhitektura Bržeg R-CNN-a je nešto kompleksnija u odnosu na prethodna dva modela, i sastoji se iz dva dela. Prvi deo čini konvoluciona neuralna mreža za određivanje predloga regija (engl. Region Proposal Network - RPN) i njen zadatak je otkrivanje objekata, dok drugi deo služi za detekciju objekata od interesa i ima strukturu kao detektor Brzog R-CNN-a. Arhitektura Bržeg R-CNN-a je prikazana na Sl. 3.



Slika 3. Arhitektura Bržeg R-CNN-a

Prilikom obučavanja modela potrebno je obučiti konvolucione mreže u oba dela i zbog toga se koriste četiri koraka naizmeničnog obučavanja. U prvom koraku se trenira

mreža za otkrivanje predloga regija. Unapred istrenirana mreža je inicijalizovana težinama dobijenim obučavanjem na ImageNet klasifikacionom setu podataka, a zatim se tokom treninga težine podešavaju tako da se što bolje detektuju regije od interesa. Tokom drugog koraka trenira se mreža iz drugog dela koja detektuje objekte. Unapred istrenirana mreža je takođe inicijalizovana težinama iz ImageNet klasifikacionog problema i tokom treninga se ove težine podešavaju, s tim da se koriste predlozi regija koje je generisala mreža za otkrivanje predloga regija. U trećem koraku težine mreže za otkrivanje objekata se inicijalizuju težinama iz drugog dela i obučavaju se samo slojevi koji su jedinstveni za prvi deo. Sada mreže iz oba dela dele parametre. U poslednjem koraku, deljeni parametri ostaju fiksni i obučavaju se samo slojevi jedinstveni za mrežu iz drugog dela.

IV. DETEKCIJA OŠTEĆENJA NA PUTU

U ovom poglavlju biće opisano funkcionisanje svakog koraka projektovanog sistema za detekciju oštećenja na putu. Na početku su učitanе slike sa pratećim anotacijama. Kako je broj parametara konvolucione neuralne mreže veliki i postoji određena kompleksnost zadatka, u trening skupu podataka se mora nalaziti proporcionalno veliki broj slika koji se daje mreži. Upravo zbog toga dolazi do potrebe za veštačkim povećanjem skupa podataka - Augmentacija slika. Augmentacija predstavlja primenu transformacija na dostupne slike kako bi se dobile nove slike i obogatio skup podataka za obučavanje mreže. Neke od transformacija koje su primenjene u radu su: Horizontalno okretanje slika, rotacija slike za 90 stepeni, zamućenje usled kretanja, zamućenje korišćenjem median filtra...

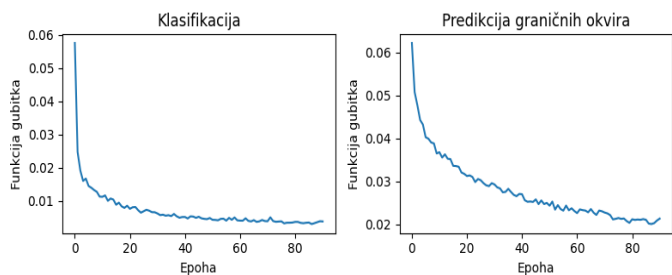
Kao što je u prethodnom poglavlju objašnjeno, kod Bržeg R-CNN-a potrebno je prvo izdvojiti mapu osobina i za to je korišćena ResNet50 konvoluciona mreža sa piramidom osobina (engl. Feature Pyramid Network - FPN). Piramida osobina se sastoji od dve putanje. Prva putanja je uobičajena konvoluciona neuralna mreža za ekstrakciju obeležja, koja sa svakim novim slojem smanjuje prostornu rezoluciju, ali se istovremeno povećava semantička vrednost tj. model "razume" šta se nalazi na slici. Iako se povećava semantičko razumevanje, gubi se informacija o prostornoj lokaciji objekata. Zbog toga se u drugoj putanji radi inverzni postupak tj. konstruišu se slojevi veće prostorne rezolucije. Konačne mape osobina se dobijaju na osnovu ove dve putanje i dodavanjem bočnih veza između njih, čime se umesto jedne generišu četiri mape osobina.

Prilikom treniranja modela računata je funkcija gubitka za klasifikaciju i predikciju graničnih okvira za oba dela modela. Mreža je optimizovana pomoću stohastičkog gradijentnog spusta sa momentumom sa hiperparametrima: konstanta učenja 0.001, momentum 0.9 i L2 parametar regularizacije 0.005.

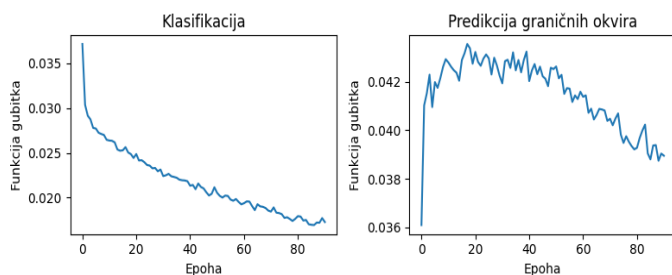
V. REZULTATI I EVALUACIJA

Komputacioni zahtevi Bržeg R-CNN-a su veoma veliki. Treniranje jedne epohe je trajalo oko 8 minuta i zbog memorijskih ograničenja, model je obučavan na 90 epoha. Na Sl. 4 su prikazane dobijene funkcije gubitka za klasifikaciju objekata i predikciju graničnih okvira kod mreže za detekciju predloga regija. Na osnovu dobijenih funkcija gubitaka vidimo

da model uspeva sve bolje sa porastom epoha da izdvaja regione od interesa kao i da sve bolje predviđa granične okvire objekata. Takođe, izračunate su funkcije gubitka za drugi deo detektora i one su prikazane na Sl. 5.

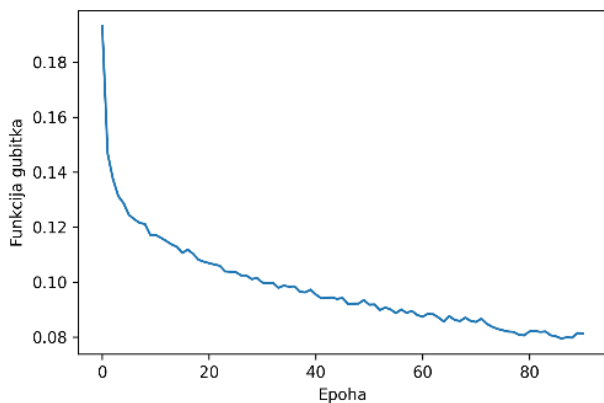


Slika 4. Funkcije gubitka za klasifikaciju regiona (levo) i predikciju graničnih okvira regiona (desno) kod mreže za detekciju predloga regiona



Slika 5. Funkcije gubitka za klasifikaciju objekata (levo) i predikciju graničnih okvira objekata (desno) kod drugog dela Bržeg R-CNN-a.

Ukupna funkcija gubitka je prikazana na Sl. 6 i ona se dobija sabiranjem prethodne četiri funkcije gubitka.



Slika 6. Ukupna funkcija gubitka Bržeg R-CNN-a.

Kada je model obučen nakon 90 epoha na trening setu podataka, potrebno je izvršiti njegovu evaluaciju i analizirati rezultate na novom setu podataka. Prosečna preciznost (engl. Average Precision - AP) i F1 skor predstavljaju najpopularnije metrike koje se koriste za evaluaciju modela kod detekcije objekata. Da bi se definisali prethodno navedeni pojmovi, neophodno je pre toga definisati jednu drugu metriku - IoU (engl. Intersection over Union). IoU je metrika u detekciji

objekata koja predstavlja stepen preklapanja između tačne pozicije objekta i predikcije. Računa se tako što se presek predikcije i tačne pozicije objekta podeli sa njihovom unijom. IoU se nalazi u opsegu između 0 i 1, gde 0 pokazuje da ne postoji preklapanje, dok 1 predstavlja savršeno preklapanje predikcije i tačne pozicije objekta. IoU je značajan jer se koristi kao prag za odluku da li je neka predikcija tačna ili ne.

IoU prag je postavljen na vrednost 0.4, pa se svaka predikcija koja ima manje preklapanje sa tačnom pozicijom objekta posmatra kao netačna detekcija, dok u suprotnom predstavlja tačnu detekciju. F1 skor se definiše kao harmonijska sredina preciznosti i odziva i računa se po formuli:

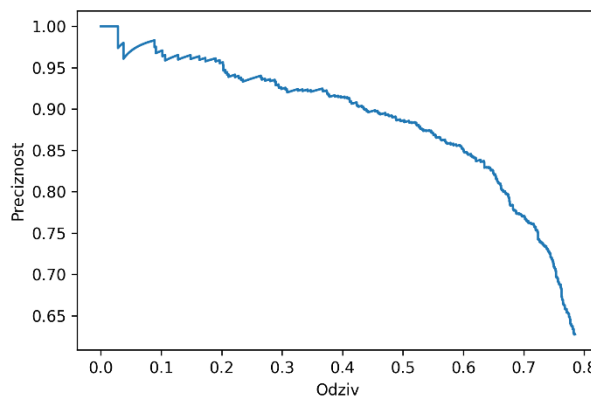
$$F_1 = 2 * \frac{Preciznost * Odziv}{Preciznost + Odziv}$$

gde preciznost predstavlja stepen tačnosti modela u identifikaciji samo relevantnih objekata i definiše se kao količnik broja tačnih predikcija i ukupnog broja predikcija, dok odziv predstavlja meru sposobnosti modela da detektuje sve tačne pozicije objekata i definiše se kao količnik tačnih detekcija i ukupnog broja objekata.

Dobijeni su sledeći rezultati:

$$Preciznost = 0.63 ; Odziv = 0.78 ; F1 skor = 0.697$$

Pored F1 skora, vrlo često se koristi i prosečna preciznost kao metrika za evaluaciju. Prosečna preciznost se računa kao površina ispod grafika Preciznost - Odziv, kod koga se na osama nalaze kumulativne vrednosti odziva i preciznosti koje se računaju za svaku predikciju počevši od onih sa najvećim indeksom pouzdanosti (verovatnoća koju daje model da je objekat dobro klasifikovan). Dobijeni grafik Preciznost-Odziv je prikazan na Sl. 7.



Slika 7. Grafik Preciznost-Odziv

Linija na grafiku Preciznost-Odziv nije monotono opadajuća već je nepravilnog oblika, ali uklanjanje ovog svojstva je moguće postići nekom od metoda interpolacije. Dobijena prosečna preciznost za interpoliran grafik iznosi 0.6989. Dobijeno vreme koje je potrebno za detekciju rupa na slici je 0.2 sekunde, što je dosta brže u odnosu na R-CNN i Brzi R-CNN, ali nedovoljno brzo da se koristi u realnom vremenu.

U nastavku su prikazani primeri kako model radi na realnim slikama, gde je sa zelenim okvirom prikazana tačna pozicija rupe, a crvenim graničnim okvirom je prikazana predikcija modela:



Slika 8. Prikaz rada modela – detektovana su sva oštećenja tačno



Slika 9. Prikaz rada modela – prepoznavanje rupa na trotoaru kao oštećenja na putu (na slici levo) i neprepoznavanje jednog objekta (na slici desno)

VI. ZAKLJUČAK

Rezultati dobijeni u ovom radu pokazuju da je moguće automatski detektovati oštećenja na putu pomoću dubokih neuralnih mreža. S obzirom da je implementiran dvostepeni detektor - Brži R-CNN, rezultati očekivano potvrđuju da je moguće jednako dobro detektovati male i velike objekte na slici. Zbog velike kompleksnosti i dugog trajanja obučavanja, model je treniran na 90 epoha, pa se može pretpostaviti da bi model za veći broj epoha postizao bolje rezultate naročito u predikciji graničnih okvira. Takođe, poboljšanje modela je moguće postići dodavanjem novih podataka koji bi pospešili sposobnost generalizacije modela. Kako je dobijeno vreme

izvršavanje na jednoj slici 0.2 sekunde, Brži R-CNN teško može da se koristi u realnom vremenu za potrebe detektovanja rupa kod autonomne vožnje, što je očekivano zbog velike kompleksnosti modela. Sa druge strane, zbog dobre tačnosti, ovako konstruisan detektor bi mogao da se iskoristi prilikom inspekcije puteva, gde bi se potpuno automatizovanim procesom detektovala oštećenja na putu i samim tim alarmirale službe za popravku oštećenja.

ZAHVALNICA

Autori zahvaljuju Fondu za inovacionu delatnost Republike Srbije koji je delom omogućio izradu ovog rada u okviru projekta broj 50301 pod nazivom "Platform for Remote development of Autonomous Driving algorithms in realistic environment - READ".

LITERATURA

- [1] Miller, Ted R., and Eduard Zaloshnja. "On a crash course: The dangers and health costs of deficient roadways." (2009).
- [2] Arya, Deeksha, et al. "Transfer learning-based road damage detection for multiple countries." *arXiv preprint arXiv:2008.13101* (2020).
- [3] Ahmed, Khaled R. "Smart pothole detection using deep learning based on dilated convolution." *Sensors* 21.24 (2021): 8406.
- [4] Girshick, Ross, et al. "Rich feature hierarchies for accurate object detection and semantic segmentation." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014.
- [5] Girshick, Ross. "Fast r-CNN." *Proceedings of the IEEE international conference on computer vision*. 2015.
- [6] Ren, Shaoqing, et al. "Faster r-CNN: Towards real-time object detection with region proposal networks." *Advances in neural information processing systems* 28 (2015).

ABSTRACT

Road infrastructure is a public good that contributes to the economic development and growth of any country, and at the same time provides various social benefits. However, the road surface deteriorates over time due to various factors such as age, traffic volume, weather conditions, engineering approach, and materials used during its construction. Therefore, knowledge about the degree of deterioration of the road is crucial for efficient and cost-effective maintenance with the aim of preserving its good and safe condition. Object detection is a complex problem because it requires the model to detect multiple, non-fixed-size objects and localize and classify them simultaneously. Also, it's challenging to make a trade-off between accuracy and training time. This paper explains and applies a two-stage detector based on deep neural networks - Faster R-CNN. The model was applied to a public dataset consisting of a few thousand potholes. The results obtained on the test data set were analyzed, demonstrating the model's ability to detect different types of road damage very well.

AUTOMATIC ROAD DAMAGE DETECTION USING DEEP LEARNING

David Đukić, Goran Ferenc, Stefan Stefanović