

Usporedba funkcija gubitka za semantičku segmentaciju objekata u prometu

G. Oreški*, S. Aničić*

* Sveučilište Jurja Dobrile u Puli/Fakultet informatike, Pula, Hrvatska
goreski@unipu.hr

Sažetak—Autonomna vožnja postaje sve važniji čimbenik svakodnevnog života, čime raste značaj dubokog učenja na kojem se takvi sustavi temelje. Posljednjih godina predložene su mnoge funkcije gubitka za problem semantičke segmentacije, čiji odabir ima direktan utjecaj na sposobnost modela da uči i uspješno generira predikcije. Ovo istraživanje analizira utjecaj različitih funkcija gubitka na performanse semantičke segmentacije na 23 različita objekta prisutna u prometu. Istraživanje je provedeno na skupu podataka prikupljenom na CARLA simulatoru prometa u urbanom okruženju. Rad obuhvaća sve najčešće korištene funkcije gubitka: *Binary Cross entropy*, *Dice*, *Focal*, *Jaccard* te *Tversky*, kako bi analizirali njihov utjecaj na konvergenciju modela. Usporedba je izvršena na tri različite arhitekture mreža: U-net, DeepLabV3Plus i FPN. Rad donosi preporuke prikladnosti pojedinih funkcija na problemu semantičke segmentacije objekata u prometu.

Ključne riječi—*semantička segmentacija, duboko učenje, autonomna vozila, funkcija gubitka*

I. UVOD

Danas sve više proizvođača automobila stavlja fokus na autonomnu vožnju, koja se brzo razvija i obećava revoluciju u životima pojedinaca. Stvaranje autonomnog sustava vožnje uključuje ne samo kontrolu i upravljanje vozilom, nego zahtijeva da vozilo i njegov autonomni sustav vožnje precizno percipiraju okolinu, očekuju i predviđaju različite pojave i događaje na cesti te donose brze i točne odluke u stvarnom vremenu, kako bi putnicima i drugim sudionicima u prometu omogućili sigurno putovanje.

U rješavanju izazova autonomne vožnje trenutno se u svijetu nameću modeli dubokog učenja. Takvi modeli su sposobni predvidjeti različita ponašanja vozila i okoline s iznimnom točnošću. Duboko učenje i računalni vid zahtijevaju korištenje skupa podataka koji je dovoljno velik i sadrži različite scenarije, tako da model može naučiti razlikovati različite obrasce i pojave na cesti. Stvaranje skupa podataka je dugotrajan i skup proces. To se pogotovo odnosi na označavanje podataka za treniranje modela s nadziranom učenjem [1]. CARLA simulator za gradsku vožnju predstavlja alternativu klasičnom načinu prikupljanju podataka. Simulator nudi niz različitih elemenata u svakodnevnom prometu, kao što su: pješaci, vozila, zgrade, ulični znakovi te drugi elementi. Također nudi mogućnost promjene doba dana, vremena, senzora, brzine, ubrzanja i drugih varijabilnih elemenata [2]. Najveća prednost takvih simulatora je mogućnost automatskog generiranja labela, čime se značajno ubrzava proces stvaranja skupa podataka.

U ovom istraživanju smo iskoristili CARLA simulator da bi istražili performanse različitih funkcija gubitka na problemu semantičke segmentacije u prometu. Funkcija gubitka je osnovni element svakog modela. Ključna je u procesu treninga te stoga njezin odabir predstavlja bitnu odluku.

II. PREGLED LITERATURE

Iako je problem autonomne vožnje danas vrlo popularan postoji malo literature vezane uz funkcije gubitka, pogotovo za semantičku segmentaciju. Jedan od razloga može biti ranije spomenuta cijena prikupljanja takvih skupova podataka. U nastavku donosimo istraživanja koja su povezana s navedenom temom.

S. Malec [3] je istražio semantičku segmentaciju za treniranje modela autonomne vožnje proširenu s podacima generiranim iz CARLA simulatora. U radu je korištena isključivo *Cross entropy* funkcija gubitka. M. Boudissa et al. [4] su korištenjem U-net modela nastojali stvoriti model za korištenje u prometu, odnosno prepoznavanje objekata te su koristili dvije funkcije gubitka: *Dice* i *Focal*. *Dice* se pokazao kao bolji izbor zbog performansi detekcije rubova objekata. O. Natan i J. Miura [5] su uz pomoć CARLA simulatora razvili model za semantičku segmentaciju i procjenu dubine. Tijekom treninga semantičke segmentacije koristili su *Dice* funkciju kombiniranu s *pixel-wise* binarnom *Cross-entropy* funkcijom u svrhu postizanja veće raznolikosti pogreške. F. Nesti et al. [6] su usporedili standardnu *Cross entropy* funkciju s vlastitom, nadograđenom funkcijom koja bolje sprječava *Adversarial Patch* napade tijekom semantičke segmentacije.

J. Ma et al. [7] su analizirali veliki broj funkcija gubitka za semantičku segmentaciju (više od 10) na medicinskim slikama. U radu se zaključuje da su najrobustnije složene funkcije gubitka između kojih je i *Dice* funkcija. M. Yeung et al. [8] su predložili novu funkciju, *Unified Focal*, koja generalizira dvije najčešće funkcije gubitka za zadatak segmentacije: *Dice* i *Cross-entropy*. Istraživanje je također provedeno na medicinskim slikama. H. Ghandorh et al. [9] su istražili način kako uz pomoć semantičke segmentacije prepoznati cestu na visokokvalitetnim satelitskim slikama, kao funkciju gubitka koristili su težinsku *Cross entropy* te *Focal Tversky* funkciju gubitka. J. Su et al. [10] su usporedili različite funkcije gubitka za nebalansiranu semantičku segmentaciju te su predložili *Cross Dropout Focal* funkciju gubitka. Koristeći istu ostvareno je 2.5%

veća točnost u usporedbi sa standardnom *Cross-entropy* funkcijom. H.X. Cheng et al. [11] su predložili novi model koji smanjuje računsku i vremensku kompleksnost semantičke segmentacije. Koristili su kombinaciju funkcija gubitka: težinski *Cross-entropy*, *Lovasz-Softmax* te *Boundary*. Kombinacijom su nastojali riješiti izazove: nebalansiranih podataka, optimizacije Intersection over Union (IoU) mjere i nejasnih granica segmentacije. S. Jadon [12] je napravio iscrpnu analizu nekoliko funkcija gubitaka te ih usporedio na primjeru semantičke segmentacije. Identificirao je u kojim situacijama je jedna prikladnija od druge. Istraživanje je proveo na medicinskim slikama.

Ovim radom donosimo sustavnu usporedbu različitih funkcija gubitka za semantičku segmentaciju na tri najpopularnije arhitekture modela. Na podacima iz prometa smo ispitali njihovu prikladnost i donijeli preporuke za korištenje. U analizu smo uključili najpopularnije funkcije spomenute u pregledu literature.

III. FUNKCIJE GUBITKA

Algoritmi dubokog učenja počivaju na mjeri koja određuje kvalitetu njihove izvedbe, odnosno koja računa veličinu pogreške između predviđenih i stvarnih (ciljnih) podataka. Funkcija gubitka predstavlja taj izračun te se koristi kao polazište za prilagođavanje parametara modela kako bi se smanjila pogreška te poboljšale performanse modela. Izbor funkcije gubitka može uvelike utjecati na performanse algoritma, samim time neprikladna funkcija će negativno utjecati na performanse i konvergenciju modela. U nastavku donosimo kratki opis funkcija obuhvaćenih ovim istraživanjem. Uz svaku funkciju dana je i njezina formula, čije detaljno objašnjenje se može pronaći u referenciranim radovima.

A. Cross Entropy

Cross entropy funkcija gubitka je definirana kao negativni logaritam predviđene vjerojatnosti za ispravnu klasu [13] te je najčešće korištena funkcija gubitka za treniranje klasifikacijskih modela. Funkcija se pokazala uspješnom na brojnim problemima te se često koristi za semantičku segmentaciju. Njezin nedostatak je što sve piksele tretira podjednako te veliki objekti dominiraju mrežom stoga su performanse na malim objektima loše. Formula kategorijske *Cross entropy* funkcije za C klasa je dana jednadžbom 1 gdje je t_i oznaka istinite klase a $f(s_i)$ *Softmax* vjerojatnost za i -tu klasu.

$$CE = - \sum_i^C t_i \log(f(s_i)) \quad (1)$$

Binary Cross entropy je varijanta navedene funkcije koja se koristi u binarnoj klasifikaciji.

$$BCE = - \sum_i^{C=2} t_i \log(f(s_i)) \quad (2)$$

B. Dice

Dice funkcija gubitka bazirana je na Sørensen–Dice koeficijentu ili Tversky indeksu [14]. Lažno pozitivni i lažno negativni primjeri su jednake važnosti te je ona imunija na problem nebalansiranosti [15]. Postoji nekoliko nedostataka, odnosno ograničenja kod treniranja neuronskih mreža s ovom funkcijom [16]: neosjetljivost na udaljenost regija koje se ne preklapaju, [17] neosjetljivost na granice objekata [18] te osjetljivost na male difuzne lezije [19]–[23]. Formula *Dice* funkcije dana je jednadžbom 3 i 4 gdje A predstavlja istinitu masku, a B predviđenu masku. Presjek maski se računa kao umnožak po elementima (engl. element-wise) te se potom sumira rezultirajuća matrica.

$$Dice = 2 * \frac{|A \cap B|}{|A| + |B|} \quad (3)$$

$$DiceLoss = 1 - Dice \quad (4)$$

C. Jaccard/Intersection over union (IoU)

Jaccard indeks, odnosno IoU gubitak daje rezultat u rasponu [0,1] za presjek između istinite i predviđene maske preko njihove unije s pretpostavkom da je 0/0 jednako 1 [24]. IoU rezultat želimo maksimizirati dok IoU gubitak želimo minimizirati tijekom treniranja. Nedostatak funkcije su loše performanse na malim objektima segmentacije. U tim slučajevima nema varijacije indeksa i ti objekti ostaju nezapaženi tijekom treninga [25]. Formula za Jaccard/IoU funkciju gubitka dana je jednadžbom 5. Oznake A i B imaju isto značenje kao u formuli 3.

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \quad (5)$$

D. Focal

Focal je varijacija *Cross entropy* funkcije. Autori su predložili dodavanje faktora $(1 - p_t)^\gamma$ standardnoj funkciji gubitka *Cross entropy* te se postavljanjem $\gamma > 0$ može smanjiti gubitak na ispravno predviđenom skupu primjera [26]. Funkcija dijeli primjere u lagane i teške te pomoću faktora povećava vrijednost gubitka za teške, odnosno smanjuje ga za lagane primjere. Funkcija gubitka klasificira svaki piksel pojedinačno što rezultira time da pogrešna anotacija značajno utječe na performanse mreže [27]. Navedeni problem nije značajan kada se koriste podaci generirani iz simulatora. Formula Focal funkcije dana je jednadžbom 6 gdje je p_t predviđena vjerojatnost ispravne klase, γ parametar koji kontrolira stupanj za koji će gubitak biti smanjen za ispravno klasificirane primjere a α težinski faktor za svaki primjer.

$$FL(p_t) = -\alpha(1 - p_t)^\gamma \log(p_t) \quad (6)$$

E. Tversky

Tversky proširuje *Dice* funkciju uvođenjem dvaju koeficijenata koji kontroliraju osjetljivost gubitka na lažno pozitivne i lažno negativne rezultate. Koeficijenti α i β

omogućuju određivanje kompromisa između preciznosti i odziva u izračunu gubitka. To rezultira fleksibilnijom funkcijom gubitka u usporedbi s gubitkom *Dice* funkcije budući da se može bolje nositi s neuravnoteženim skupovima podataka i smanjiti utjecaj klasa s velikim brojem piksela [28]. Tversky funkcija gubitka je dana jednadžbom 7 gdje TP označava istinito pozitivne piksele, FP lažno pozitivne, FN lažno negativne a α i β su ranije spomenuti koeficijenti.

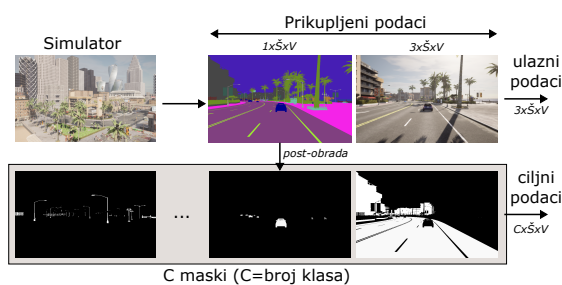
$$T_{\alpha,\beta}(y, \hat{y}) = \frac{TP}{TP + \alpha FP + \beta FN} \quad (7)$$

IV. EKSPERIMENT I REZULTATI ANALIZE

Da bi dali sustavan prikaz utjecaja funkcija gubitka na semantičku segmentaciju prometa, u istraživanje su uključene tri arhitekture najčešće korištene u literaturi što ukupno daje 15 različitih modela koje smo vrednovali koristeći četiri mjere uspješnosti predikcije: IoU, F mjera, preciznost i odziv. Detaljan opis eksperimenta zajedno s rezultatima je dan u nastavku.

A. Skup podataka iz simulatora CARLA

Skup podataka je prikupljen koristeći CARLA simulator autonomne vožnje. Za generiranje podataka kreirana su dva različita senzora, jedan za generiranje RGB slike, a drugi za generiranje semantičke segmentacije. Trenutačna verzija CARLA-e sadržava 23 različite klase ($C=23$). Senzor za semantičku segmentaciju označava klase svakog piksela oznakama od 1 do 23 u crvenom (R) kanalu slike, dok su ostala dva kanala (G i B) prazna, odnosno sadržavaju vrijednost 0. Koristeći funkciju *cityScapesPalette* ta slika se može transformirati iz slike s vrijednosti samo u crvenom kanalu u sliku s vrijednostima u svim kanalima te se grafički lijepo prikazati. Za generiranje prometa koristila se skripta *generate_traffic.py*, a za postavljanje Ego vozila (vozila s kamerom), pokretanje senzora te autopilot, odvojena, vlastita skripta.



Slika 1: Generiranje skupa podataka

Slika 1 prikazuje dohvaćanje podataka iz simulatora. Podaci sa senzora se dohvaćaju u paru te sadržavaju RGB sliku i masku u jednom kanalu (koja se iscrtava kao RGB sa spomenutom *cityScapesPalette* funkcijom). Slika je veličine 3 x Širina(\hat{S}) x Visina(V), a maska 1x \hat{S} x V . U post-obradi tj. pripremi ciljnih podataka, maska se tehnikom *one-hot encoding* dijeli na C zasebnih kanala u kojima se nalaze odvojene maske svake klase. Na Slici

1 su odvojene maske vidljive kao crno-bijele slike gdje bijeli pikseli prikazuju prisutnost pojedine klase. Modelu se tijekom treniranja kao ulazni podatak daje slika i maska veličine $C \times \hat{S} \times V$. Model predviđa masku za svaku pojedinu klasu na zasebnom kanalu. Da bi se nakon predikcije došlo do inicijalne dimenzije maske ($1 \times \hat{S} \times V$) izlaz se agregira koristeći *argmax* funkciju po prvoj dimenziji. Prikupljeno je 11000 slika, koje su raspodijeljene u omjeru 70/20/10 u trening, validacijski i testni skup respektivno.

B. Modeli

Ovaj rad analizira pet različitih funkcija gubitka koje su korištene za trening u kombinaciji s tri različite arhitekture mreže. Kao osnovna mreža (*engl. backbone*) odabran je *resnext50*, a korištene težine su pred-trenirane na *imagenet* skupu podataka. Za arhitekture smo koristili: FPN (*Feature Pyramid Network*) [29], DeepLabV3+ [30] te U-net [31].

FPN arhitektura se koristi za izdvajanje značajki slike (*engl. feature extractor*). Kao ulaz prima sliku proizvoljne veličine i generira mape značajki proporcionalne veličine na više razina. DeepLabV3Plus koristi modul *atrous spatial pyramid pooling* (ASPP) za izdvajanje kontekstualnih informacija na više razina (*engl. multi-scale*) iz mapa značajki konvolucijske neuronske mreže. U-net je enkoder-dekoder tip arhitekture koji prvo uz pomoć enkodera izdvaja značajke te potom dekoderom rekonstruira segmentacijsku mapu.

C. Evaluacijske mjere

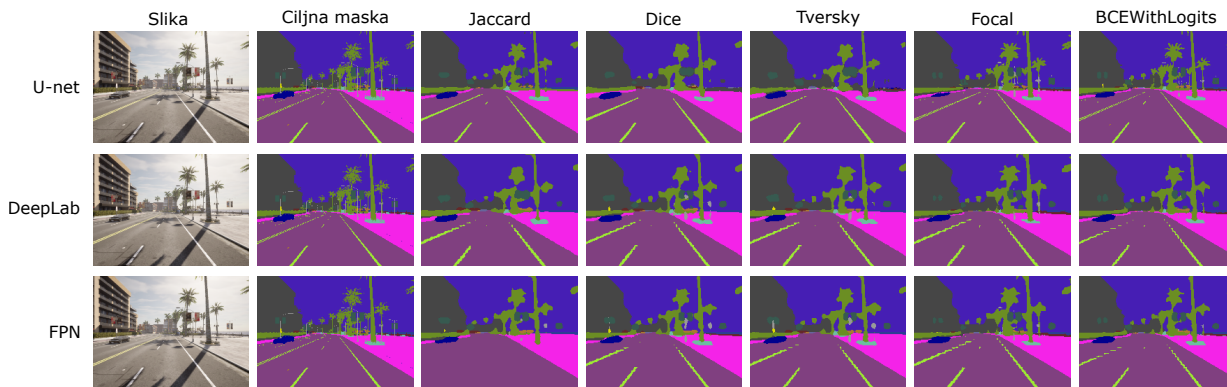
Evaluacijske mjere se koriste za procjenu uspješnosti modela, a mi ih koristimo za usporedbu prikladnosti odabranih funkcija gubitka. Mnoge mjere imaju polazište u matrici konfuzije (*engl. confusion matrix*). U matrici postoje 4 moguća ishoda predikcije: istinito pozitivna, istinito negativna, lažno pozitivna te lažno negativna. Različite mjere predstavljaju različite odnose tih ishoda. U nastavku su opisane korištene evaluacijske mjere u radu.

Preciznost je udio točno klasificiranih piksela u skupu pozitivno klasificiranih piksela. Odziv je udio točno klasificiranih piksela u skupu svih pozitivnih piksela. F-mjera je harmonijska sredina preciznosti i odziva. IoU mjera je udio ispravno klasificiranih piksela klase u skupu svih piksela koji toj klasi pripadaju ili su svrstani u nju.

D. Rezultati

Trening modela je zaustavljan kada je vrijednost funkcije gubitka prestala padati na validacijskom skupu podataka. Konačni rezultati uspješnosti modela na testnom skupu podataka, iskazani ranije spomenutim mjerama, su vidljivi u Tablici I. Na Slici 2 je dan usporedni prikaz predikcije modela za isti ulazni podatak. Analizirane su performanse pojedinih modela kroz funkcije gubitka, a potom i korištenu arhitekturu.

Iz Tablice I je vidljivo da je *Binary Cross entropy* funkcija ostvarila najbolje rezultate mjereno F mjerom, IoU mjerom i preciznosti predikcije. U Poglavlju III je



Slika 2: Prikaz predikcija svih modela na istoj slici

istaknuto da je to najčešće korištena funkcija te da ostvaruje dobre rezultate na semantičkoj segmentaciji što ovo istraživanje potvrđuje. Štoviše, iz Slike 2 je vidljivo da su i manji objekti prepoznati, što se često navodi kao glavni nedostatak te funkcije. Rezultati pokazuju da je *Binary Cross entropy* robusna funkcija te da se može koristiti s bilo kojom arhitekturom.

Focal funkcija gubitka ima vrlo slične rezultate kao i *Binary Cross entropy*, što dokazuje da se radi o njezinoj varijanti koja dodaje jedan termin. Obzirom da je *Focal* funkcija parametrizirana s γ , kojim se određuje koliko će biti smanjen utjecaj laganih primjera na vrijednost funkcije, njegovim mijenjanjem se može utjecati na konačan ishod. U eksperimentu smo zadržali zadanu vrijednost $\gamma=2$ koju autori navode kao najbolji odabir. *Focal* funkcija je konzistentno kroz sve arhitekture imala nešto bolju preciznost i manji odziv od *Cross entropy* funkcije. Ako usporedimo dvije funkcije na Slici 2, vidimo da udaljene linije između prometnih traka (manji objekti) bolje prepoznaje *Focal* funkcija, dok automobil bolje prepoznaje *Cross entropy*.

Dice funkcija i njezino proširenje *Tversky* ostvaruju slične rezultate, koji su lošiji od ranije spomenutih tehnika. Ukoliko se koristi *U-net* ili *FPN* tada *Dice* postiže bolje rezultate, dok s *DeepLabV3Plus* je obrnuto. Najveća razlika je vidljiva na *U-net* mreži gdje *Dice* ostvaruje značajno bolji rezultat. Za razliku od prošle dvije tehnike, ove obje imaju veći odziv nego preciznost, bez obzira na korištenu mrežu. Vizualnom analizom predikcija sa Slike 2 je vidljivo da obje funkcije postižu lošije rezultate od *Cross entropy* i *Focal* funkcije, iako na nekim elementima kao što je stup rasvjete s lijeve strane ceste (žuta boja na maski) imaju bolju detekciju.

Jaccard funkcija postiže dobre rezultate ako se ne koristi s *FPN* mrežom u kojem slučaj rezultira s najlošijim modelom. Iznimno je dobra u kombinaciji s *U-net* arhitekturom jer postiže rezultate tek nešto lošije od *Focal* funkcije. Međutim, za razliku od nje ima bolje vrijednosti odziva. Na slici je vidljivo da *Jaccard* funkcija najlošije prepoznaje male i udaljene objekte kao što su znakovi, stupovi i linije prometnih traka.

Tablica I: Rezultati usporedbe funkcija gubitka

Arhitektura mreže	Funkcija gubitka	Mjera _{rest}			
		F mjera	IOU	Preciznost	Odziv
U-net	Jaccard	0.8836	0.7943	0.8352	0.9385
	Dice	0.8805	0.7889	0.8387	0.9272
	Tversky	0.849	0.74	0.7781	0.935
	Focal	0.9078	0.8351	0.9552	0.8653
	BCEWithLogits	0.9179	0.8524	0.9386	0.8983
DeepLabV3Plus	Jaccard	0.8205	0.6974	0.7192	0.9562
	Dice	0.7558	0.6086	0.6224	0.9637
	Tversky	0.762	0.6172	0.6319	0.9619
	Focal	0.8914	0.8076	0.9591	0.8334
	BCEWithLogits	0.908	0.835	0.9371	0.8809
FPN	Jaccard	0.5535	0.3833	0.3907	0.9504
	Dice	0.6142	0.4437	0.4479	0.9784
	Tversky	0.6021	0.4313	0.4352	0.9784
	Focal	0.8903	0.8056	0.9617	0.8296
	BCEWithLogits	0.9081	0.8351	0.9389	0.8795

Analizom rezultata iz Tablice I jasno se vidi da je vrlo bitno voditi brigu koju funkciju gubitka koristiti s kojom arhitekturom jer rezultati mogu biti značajno različiti. *U-net* mreža je ostvarila najbolje rezultate. Sve funkcije u kombinaciji s njom postižu najbolje rezultate F i IoU mjere. Slike koje prikazuju kombinaciju *U-net* i *Cross entropy* ili *Focal* funkcije su iznimno dobre te se na njima vide detalji kao što su grane i lišće na drvećima, što nije prisutno na slikama ostalih modela. S druge strane *FPN* nije dobro kombinirati s *Jaccard*, *Dice* i *Tversky* funkcijama jer generirani modeli imaju značajno lošije performanse.

V. ZAKLJUČAK

Cilj ovog rada je bio usporediti najpopularnije funkcije gubitka za semantičku segmentaciju na skupu slika iz prometa. U tu svrhu smo generirali slike i njihove anotacije pomoću CARLA simulatora. U eksperiment smo uključili tri različite arhitekture mreža koje su najzastupljenije u dostupnoj literaturi.

Nakon provedenog istraživanja možemo zaključiti da je *Binary Cross entropy* funkcija postigla najbolje rezultate bez obzira na odabranu arhitekturu. Dovoljno je robusna da bi ju svakako trebalo razmotriti u inicijalnom koraku stvaranja modela semantičke segmentacije. Najbolja se pokazala njezina kombinacija s *U-net* arhitekturom koja je ostvarila vrlo uvjerljive rezultate i impresivne slike maski. Osim *Binary Cross entropy* funkcije i *Focal* funkcija je dobar odabir jer otvara prostor da pomoću validacije γ parametra dođemo do još boljih rezultata. Ukoliko je odziv

bitna mjera tada je preporučena kombinacija *U-net* mreže i *Jaccard* funkcije.

LITERATURA

- [1] Q. Wang, Y. Ma, K. Zhao, and Y. Tian, "A comprehensive survey of loss functions in machine learning," *Annals of Data Science*, pp. 1–26, 2020.
- [2] W. Yuan and W. Xu, "Gaploss: A loss function for semantic segmentation of roads in remote sensing images," *Remote Sensing*, vol. 14, no. 10, p. 2422, 2022.
- [3] S. Malec, "Semantic segmentation with carla simulator," 2021.
- [4] M. Boudissa, H. Kawanaka, and T. Wakabayashi, "semantic segmentation of traffic landmarks using classical computer vision and u-net model," in *Journal of Physics: Conference Series*, vol. 2319, no. 1. IOP Publishing, 2022, p. 012031.
- [5] O. Natan and J. Miura, "Semantic segmentation and depth estimation with rgb and dvs sensor fusion for multi-view driving perception," in *Pattern Recognition: 6th Asian Conference, ACPR 2021, Jeju Island, South Korea, November 9–12, 2021, Revised Selected Papers, Part I*. Springer, 2022, pp. 352–365.
- [6] F. Nesti, G. Rossolini, S. Nair, A. Biondi, and G. Buttazzo, "Evaluating the robustness of semantic segmentation for autonomous driving against real-world adversarial patch attacks," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2022, pp. 2280–2289.
- [7] J. Ma, J. Chen, M. Ng, R. Huang, Y. Li, C. Li, X. Yang, and A. L. Martel, "Loss odyssey in medical image segmentation," *Medical Image Analysis*, vol. 71, p. 102035, 2021.
- [8] M. Yeung, E. Sala, C.-B. Schönlieb, and L. Rundo, "Unified focal loss: Generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation," *Computerized Medical Imaging and Graphics*, vol. 95, p. 102026, 2022.
- [9] H. Ghandorh, W. Boulila, S. Masood, A. Koubaa, F. Ahmed, and J. Ahmad, "Semantic segmentation and edge detection—approach to road detection in very high resolution satellite images," *Remote Sensing*, vol. 14, no. 3, p. 613, 2022.
- [10] J. Su, S. Anderson, and L. S. Mihaylova, "A deep learning method with cross dropout focal loss function for imbalanced semantic segmentation," in *2022 Sensor Data Fusion: Trends, Solutions, Applications (SDF)*. IEEE, 2022, pp. 1–6.
- [11] H.-X. Cheng, X.-F. Han, and G.-Q. Xiao, "Cenet: Toward concise and efficient lidar semantic segmentation for autonomous driving," in *2022 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2022, pp. 01–06.
- [12] S. Jadon, "A survey of loss functions for semantic segmentation," in *2020 IEEE conference on computational intelligence in bioinformatics and computational biology (CIBCB)*. IEEE, 2020, pp. 1–7.
- [13] L. Li, M. Doroslovački, and M. H. Loew, "Approximating the gradient of cross-entropy loss function," *IEEE Access*, vol. 8, pp. 111 626–111 635, 2020.
- [14] H. Kervadec, J. Bouchtiba, C. Desrosiers, E. Granger, J. Dolz, and I. B. Ayed, "Boundary loss for highly unbalanced segmentation," in *International conference on medical imaging with deep learning*. PMLR, 2019, pp. 285–296.
- [15] X. Li, X. Sun, Y. Meng, J. Liang, F. Wu, and J. Li, "Dice loss for data-imbalanced nlp tasks," *arXiv preprint arXiv:1911.02855*, 2019.
- [16] Y. Zhang, S. Liu, C. Li, and J. Wang, "Rethinking the dice loss for deep learning lesion segmentation in medical images," *Journal of Shanghai Jiaotong University (Science)*, vol. 26, pp. 93–102, 2021.
- [17] H. Rezatofghi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 658–666.
- [18] J. Dolz, C. Desrosiers, and I. B. Ayed, "3d fully convolutional networks for subcortical segmentation in mri: A large-scale study," *NeuroImage*, vol. 170, pp. 456–470, 2018.
- [19] N. Guizard, P. Coupé, V. S. Fonov, J. V. Manjón, D. L. Arnold, and D. L. Collins, "Rotation-invariant multi-contrast non-local means for ms lesion segmentation," *NeuroImage: Clinical*, vol. 8, pp. 376–389, 2015.
- [20] R. Harmouche, N. K. Subbanna, D. L. Collins, D. L. Arnold, and T. Arbel, "Probabilistic multiple sclerosis lesion classification based on modeling regional intensity variability and local neighborhood information," *IEEE transactions on Biomedical Engineering*, vol. 62, no. 5, pp. 1281–1292, 2014.
- [21] M. Styner, J. Lee, B. Chin, M. Chin, O. Commowick, H. Tran, S. Markovic-Plese, V. Jewells, and S. Warfield, "3d segmentation in the clinic: A grand challenge ii: Ms lesion segmentation," *MIDAS journal*, vol. 2008, pp. 1–6, 2008.
- [22] K. C. Wong, M. Moradi, H. Tang, and T. Syeda-Mahmood, "3d segmentation with exponential logarithmic loss for highly unbalanced object sizes," in *Medical Image Computing and Computer Assisted Intervention—MICCAI 2018: 21st International Conference, Granada, Spain, September 16–20, 2018, Proceedings, Part III 11*. Springer, 2018, pp. 612–619.
- [23] C. Lucas, A. Kemmling, A. M. Mamlouk, and M. P. Heinrich, "Multi-scale neural network for automatic segmentation of ischemic strokes on acute perfusion images," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE, 2018, pp. 1118–1121.
- [24] M. Berman, A. R. Triki, and M. B. Blaschko, "The lovasz-softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4413–4421.
- [25] U. Kamal, T. I. Tonmoy, S. Das, and M. K. Hasan, "Automatic traffic sign detection and recognition using segunet and a modified tversky loss function with l1-constraint," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 4, pp. 1467–1479, 2019.
- [26] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.
- [27] L. Wang, C. Wang, Z. Sun, and S. Chen, "An improved dice loss for pneumothorax segmentation by mining the information of negative areas," *IEEE Access*, vol. 8, pp. 167 939–167 949, 2020.
- [28] H. Gábor, "The tversky loss function and its modifications for medical image segmentation," 2022.
- [29] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2117–2125.
- [30] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 801–818.
- [31] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18*. Springer, 2015, pp. 234–241.