

DETEKCIJA I SEGMENTACIJA ZUBA NA PANORAMSKIM RADIOGRAFSKIM SNIMCIMA KORIŠĆENJEM MAŠINSKOG UČENJA**DETECTION AND SEGMENTATION OF TEETH IN PANORAMIC RADIOGRAPHIC IMAGES USING MACHINE LEARNING**

Dorina Petrović, *Fakultet tehničkih nauka, Novi Sad*

Oblast – ELEKTROTEHNIKA I RAČUNARSTVO

Kratak sadržaj – Panoramski radiografski snimci koriste se u stomatologiji kao pomoćni alat pri prevenciji i dijagnostici zubnih oboljenja. U radu je predložen jedan postupak za automatizovanu detekciju i segmentaciju pojedinačnih zuba na panoramskim radiografskim snimcima vilice. Metod se zasniva na primeni Mask R-CNN algoritma i metodologiji za anotaciju radiografskih snimaka zuba. Razmatrane su strategije mašinskog učenja i performanse sistema zasnovanih na skupovima za obuku različitih veličina. Poređene su karakteristike dobijenih rešenja i potrebnog vremena obuke nad četiri skupa podataka različitih veličina.

Ključne reči: segmentacija zuba, Mask R-CNN, panoramski radiografski snimci, kompjuterska vizija

Abstract – Panoramic radiographs are used in dentistry as an auxiliary tool in the prevention and diagnosis of dental diseases. The paper proposes a procedure for automated detection and segmentation of individual teeth on panoramic radiographic images of the jaw. The method is based on the application of the Mask R-CNN algorithm and the methodology for annotation of dental radiographs. Machine learning strategies and performance of systems based on training sets of different sizes are discussed. The characteristics of the obtained solutions and the required training time on four datasets of different sizes were compared.

Keywords: tooth segmentation, Mask R-CNN, panoramic radiographic images, computer vision

1. UVOD

Panoramski radiografski snimci koriste se u stomatologiji kao pomoćni alat pri prevenciji i dijagnostici zubnih oboljenja. Rendgen snimci omogućavaju prikaz strukture i anomalija zuba, pregled vilice, ukazuju na prisustvo cisti, tumora itd. Stručnjaci neretko na osnovu kvaliteta snimka, kao i ličnog znanja i tumačenja donose zaključak o daljem toku terapije. Tako, svako lečenje koje je usko povezano sa radiografskim snimcima vilice, zavisi od kvaliteta istih, ali i od stručnog znanja stomatologa. Usled prisustva šuma u rendgenskim snimcima, može se dešavati da dođe do pogrešnog tumačenja, što dovodi do komplikacija daljeg toka lečenja.

NAPOMENA:

Ovaj rad proistekao je iz master rada čiji mentor je bio dr Branko Brkljač, vanr. prof.

Sve širom upotrebom veštačke inteligencije [1] pojavila se ideja za korišćenjem dubokih neuronskih mreža za potrebe detekcije zuba, karijesa, kao i raznih anomalija na panoramskim radiografskim snimcima vilice. Postojanje sistema za automatsko detektovanje i segmentaciju zuba omogućilo bi smanjenje grešaka koje su uzrokovane umorom, nepažnjom i drugim faktorima. Ovim radom predlaže se rešenje problema segmentacije pojedinačnih zuba na panoramskim radiografskim snimcima vilica, korišćenjem Mask-RCNN algoritma za segmentaciju objekata u slici.

2. PREDLOŽENO REŠENJE

Cilj predložene eksperimentalne postavke bio je pronalaženje načina za efikasno doobučavanje unapred trenirane Mask R-CNN mreže za potrebe formiranja modela koji omogućava segmentaciju pojedinačnih zuba na panoramskim radiografskim snimcima vilice, odnosno automatizovano izdvajanje njihovih kontura. Primer željenog rezultata prikazan je na Slici 1, koja pored kontura pojedinačnih zuba ilustruje i činjenicu da usled usvojene tehnike snimanja postoji i inherentno preklapanje anatomskih struktura na panoramskom snimku zuba, što u značajno meri otežava njihovu segmentaciju (izdvajanje kontura).



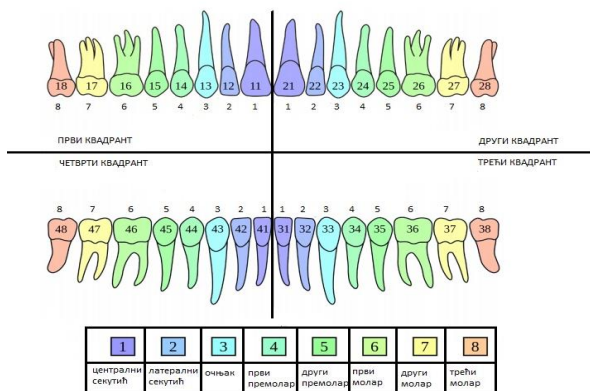
Slika 1. Izgled ručno anotirane slike sa polinomskom reprezentacijom kontura zuba (uočiti preklapanja)

S obzirom da se inicijalni skup podataka kreiran na početku istraživanja sastojao od 20 anotiranih radiografskih snimaka, pokazalo se da pomenuta baza predstavlja relativno mali skup podataka za nezavisno obučavanje mreže koja sadrži veliki broj parametara, kao što je Mask R-CNN. Da bi se nadomestio opisani nedostatak nedovoljne veličine kreirane baze snimaka, pošlo se od ideje da se umesto potpune i nezavisne obuke mreže koja podrazumeva određivanja parametara modela polazeći od slučajnih inicijalnih vrednosti težinskih koeficijenata, kao polazne vrednosti parametara koriste

vrednosti težinskih koeficijenata mreže koja je već ranije obučavana na MSCOCO skupu podataka [2]. Iako je ovakav pristup već praktikovan u literaturi, dodatni izazov predstavljala je činjenica da su slike i kategorije objekata korišćene za obuku originalnog modela potpuno drugačijeg modaliteta od razmatranih radiografskih snimaka i zuba, kao objekta od interesa. Međutim, imajući u vidu da je predložena duboka neuronska mreža obučena na MSCOCO skupu podataka nelinearna struktura za hijerarhijsko izdvajanje vizuelnih obeležja, pretpostavljeno je da se zamenom završnih slojeva mreže sa novim parametrima preostali elementi mreže mogu iskoristiti kao način za efikasno izdvajanje diskriminativnih informacija o objektima od interesa.

2.1. Skup podataka i proces anotiranja

Primarni skup podataka sastojao se od 20 panoramskih radiografskih snimaka, anotiranih pomoću VIA softvera za manuelnu anotaciju slika [3]. Primer anotiranih snimaka na kojima su ucrtane konture pojedinačnih zuba prikazan je na Slici 1. Svaki od korišćenih snimaka u skupu za obuku predstavlja sivu sliku zapisano u .jpeg formatu. Prilikom anotiranja zuba poštovan je standard označavanja prikazan na Slici 2.



Slika 2. FDI standard obeležavanja zuba. Svaki zub je obeležen dvocifrenim brojem, gde prva cifra predstavlja kvadrant kom zub pripada, a druga cifra tip zuba (u tabeli ispod). Prikaz je baziran na ilustraciji iz [4].

2.2. Arhitektura Mask R-CNN

Mask R-CNN predstavlja unapređenje Faster R-CNN arhitekture. U osnovi ovog sistema nalazi se ResNet-101 mreža koja je zadužena za ekstrakciju obeležja iz slika. Na izlazu ove mreže generiše se mapa obeležja, koja predstavlja ulaz u engl. Region Proposal Network, RPN, čija uloga je da skenira dobijenu mapu obeležja i predlaže regione od interesa (engl. Region of Interest, RoI) koji potencijalno sadrže objekte. U narednom regionu od interesa se poravnavaju sa stvarnom pozicijom objekta (engl. RoIAlign). Na krajnjem izlazu mreže, u paraleli se pojavljuju: I) Potpuno povezani sloj (engl. Fully Connected Layer), koji je zadužen za generisanje koordinata prozora u kome se nalazi objekat i odgovarajuće klase objekta (u posmatranom slučaju binarne: 0-pozadina i 1-zub); kao i izlaz II) Konvoluciona mreža koja generiše odgovarajuće binarne maske.

2.3. Opis eksperimentalne postavke

S obzirom na to da se naš inicijalni skup podataka sastoji od 20 anotiranih radiografskih snimaka, ovo predstavlja

relativno mali skup podataka za obučavanje mreže koja sadrži veliki broj parametara, kao što je Mask R-CNN. Da bi se nadomestio nedostatak podataka, koristiće se vrednosti težinskih koeficijenata modela neuronske mreže koji je obučavan na MSCOCO skupu za obuku. Pored opisanog eksperimenta koji je podrazumevao optimizaciju parametara modifikovane mreže na inicijalnom skupu podataka od 20 slika, u radu su razmatrane još tri eksperimentalne postavke korišćenjem *OdontoAI* baze slika predložene u [4]. Tako su analizirane performanse modela doobučavanih na skupovima podataka preuzetim iz *OdontoAI* baze snimaka, tj. skupovima za obuku od 100, 1000 i 2000 anotiranih slika. Eksperimenti su imali za cilj da utvrde optimalni broj slika koje je potrebno obezbediti za obuku modela da bi se postigla što uspešnija i preciznija detekcija, odnosno segmentacija zuba, a uz što manje utrošenih resursa (trajanja obuke i broja anotiranih snimaka). U svim slučajevima, eksperiment je postavljen tako da se prilikom doobučavanja modela najpre treniraju samo završni slojevi mreže u trajanju od 15 epoha nad datim skupom podataka, nakon čega se preostali parametri mreže (preuzeti iz usvojenog predobučanog modela) najpre „odmrznu“, a zatim u narednih, dodatnih 15 epoha, podešavaju prilikom doobučavanja čitave mreže. Na taj način, optimizacija parametara je u svim slučajevima razmatrana u trajanju od ukupno 30 epoha. Nad skupom podataka od 20 slika, evaluacija performansi je realizovana korišćenjem unakrsne validacije sa jednim izdvojenim elementom, odnosno kao prosečna vrednost 20 ponovljenih eksperimenata. U slučaju preostala tri eksperimenta, skupovi podataka su formirani na slučajnom način izborom snimaka iz *OdontoAI*, odnosno slučajnom podelom na skup za obuku i validaciju po principu 80-20. Za postupak optimizacije, pri optimizaciji mreže korišćen je algoritam stohastičkog opadanja gradijenta (engl. Stochastic Gradient Descent, SGD). U Tabeli 1. nalaze se informacije o hiperparametrima: veličini podskupa za obuku u jednoj iteraciji opadanja gradijenta, i usvojenoj brzini učenja za odgovarajući skup podataka (broj uzoraka). Model je analiziran korišćenjem NVIDIA RTX 4090 grafičke kartice i i9 10980XE procesora.

2.3. Način merenja uspešnosti rada sistema

Za izračunavanje uspešnosti rada sistema korišćena je usrednjena prosečna preciznost (engl. mean Average Precision, mAP), prosečna preciznost (engl. precision) i odziv ili osetljivost (engl. recall) detektora, kao i vizuelna procena kvaliteta segmentacije zuba od pozadine unutar prozora koji sadrži objekat. Agregatna mAP metrika ukazuje na generalnu sposobnost modela da detektuje i lokalizuje objekte u slici za sve razmatrane vrednosti praga odlučivanja, najčešće u opsegu od 0.5 do 0.95. Kao i preostale metrike, zasniva se statistikama koje procenjuju broj uspešno detektovanih objekata, a na osnovu kriterijuma odlučivanja koji razmatra odnos preseka i unije površina stvarnog i detektovanog prozora (engl. Intersection over Union, IoU). Npr. za vrednost praga odlučivanja 0.5, ukoliko je $IoU > 0.5$, detekcija se smatra stvarnim pozitivom (engl. true positive, TP). Mera je zasnovana na stepenu preklapanja između predikovanog prozora (engl. predicted bounding box) unutar koga se nalazi objekat i njegove ispravno označene vrednosti (engl. ground-truth bounding box). Slično, ako

za meru važi da je $IoU < 0.5$, detekcija se smatra lažnim pozitivom (engl. *False positive*, FP). Sa druge strane, ako je presek 0, onda se radi o promašaju, odnosno lažnom negativu (engl. *false negative*, FN). Stvarni negativ (engl. *true negative*, TN) se ne evaluiraju, jer se pretpostavlja da svaka slika iz test skupa ima objekat u sebi.

Tako se preciznost detekcije definiše kao odnos stvarnih pozitivna (TP) i ukupnog broja detekcija (TP+FP), a osetljivost ili odziv kao odnos broja stvarnih pozitivna (TP) i ukupnog broja objekata u slici (TP+FN). Pri svakoj vrednosti praga, uprosečavanjem po svim slikama iz test skupa, dobija se par vrednosti koji označava prosečnu preciznosti i odziv sistema, odnosno radnu karakteristiku detektora. Na osnovu dobijene empirijske krive, mAP vrednost se dobija kao interpolirana površina ispod krive. Sa druge strane, pre ocenjivanja uspešnosti generisanih detekcija, za potiskivanje nepouzdatih odluka i višestrukih detekcija uglavnom se koriste vrednosti uslovne verovatnoće prisutnosti objekta u generisanom prozoru i stepen preklapanja okolnih prozora sa prozorom kome lokalno odgovara najpouzdanija odluka, tj. sa prozorom detekcije koja je lokalno najpouzdanija (engl. non-maxima suppression, NMS).

3. REZULTATI I DISKUSIJA

Za doobučavanje skupa podataka sa inicijalnih 20 slika bilo je potrebno oko 45 minuta. Pri evaluaciji, dobijena mAP vrednost, uprosečena po svim slikama, tj. eksperimentima, bila je 26%. Ovakva vrednost predstavlja lošiji rezultat u odnosu na eksperimente sa preostalim skupovima podataka prikazane u Tabeli 1, što ukazuje da je veličina ovakvog skupa podataka nedovoljna da bi model imao potreban nivo preciznosti.

Tabela 1. Karakteristike detektora dobijenih doobučavanjem usvojenog modela nad odgovarajućim skupovima podataka sa 100, 1000 i 2000 snimaka

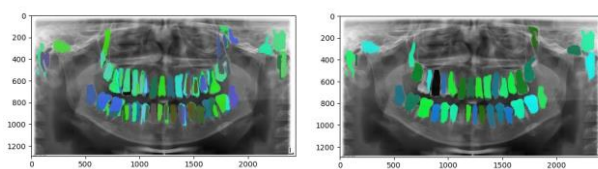
	ПОДСКУП УЗОРАКА ЗА ОБУКУ (engl. Batch size)	БРИЗНА УЧЕЊА (engl. Learning rate) завршни слојеви/сви слојеви	мАР (engl. mean Average Precision)	ПРЕЦИЗНОСТ (engl. Precision)	ОСЕЋЉИВОСТ (engl. Recall)	ТРАЈАЊЕ ОБУКЕ (engl. Training time) при тренирању завршних слојева + свих слојева
Скуп података са 100 узорака (слика)	2	0.05/0.01	75 %	98 %	80 %	~ 60 МИН
Скуп података са 1000 узорака (слика)	2	0.001/0.0001	85%	98.8%	88.5%	~ 280 МИН
Скуп података са 2000 узорака (слика)	4	0.05/0.01	86 %	99 %	90%	~ 600 МИН

Iako skup podataka sa najvećim brojem slika očekivano daje najbolje karakteristike detektora, takav skup za obuku je zahtevan za anotaciju, a ujedno i vremenski najzahtevniji po pitanju trajanja obuke. Tokom eksperimenata korišćene su različite brzine učenja. Primarna brzina učenja izabrana je empirijski, za potrebe treniranja modela sa 100 slika, pri čemu su brzine učenja prilagođavane predloženim fazama obuke modela: optimizaciji završnih slojeva mreže, i u drugoj fazi optimizaciji svih slojeva, Tabela 1. Završni ili izlazni slojevi mreže, tj. modela, obučavani su polazeći od slučajnih vrednosti težinskih koeficijenata kako bi se postiglo optimalno odlučivanje (detekcija i segmentacija) na osnovu obeležja izdvojenih korišćenjem ranije naučenih deskriptora slike. Nakon toga, kako bi se

dobitno poboljšale performanse, čitava mreža je obučavana kroz još jedan ciklus od 15 epoha, kako bi se postiglo fino podešavanje i ujednačavanje preuzetih vrednosti parametara sa parametrima završnih slojeva mreže. Praćenjem promena vrednosti ciljne funkcije, tj. greške obuke, utvrđeno je da je najbolja brzina učenja nad skupom podataka sa 100 snimaka 0.01.

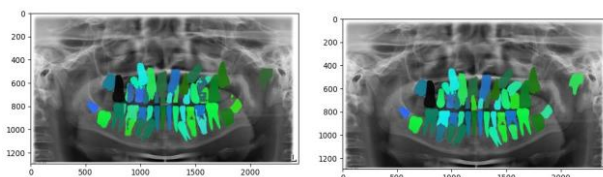
Za dobijanje opadajuće krive pri treniranju skupova sa većim brojem snimaka bilo je neophodno smanjiti brzinu učenja. Zbog ograničenih resursa nad skupom od 2000 slika nije testirano ponašanje za slučaj treniranja sa manjom brzinom učenja, što bi bilo zanimljivo posmatrati. Izbor veličine podskupa uzoraka za obuku (engl. batch size) tokom jedne iteracije opadanja gradijenta biran je po ugledu na originalni rad [5]. Povećanjem veličine podskupa uzoraka za obuku, nisu dobijene bolje vrednosti, a ni bolja brzina.

Na slikama 3, 4, 6 i 8 prikazani su vizuelni rezultati detekcije i segmentacije zuba, dok su na slikama 5, 7, i 9 prikazane procenjene vrednosti prosečne preciznosti, odziva i usrednjene prosečne preciznosti detektora. Snimci korišćeni za vizuelnu interpretaciju kvaliteta segmentacije nisu bili deo skupova za obuku, niti validaciju modela.

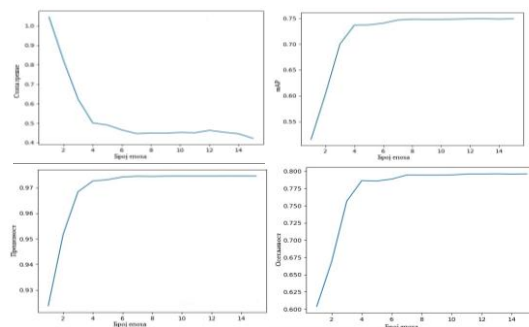


Slika 3. Segmentacija nakon doobučavanja nad skupom od 20 snimaka (desno, nakon NMS)

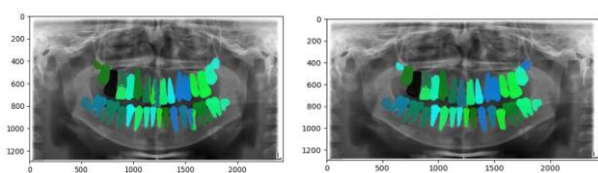
Nad svakim prikazanim snimkom primenjen je algoritam supresije detekcija koje ne odgovaraju maksimumu (NMS), odnosno maske čije su pouzdanosti ispod praga pouzdanosti od 0.3 su odbačene. Takođe, za NMS kao i prikazane dijagrame preciznosti i odziva korišćena je vrednost IoU od 0.5.



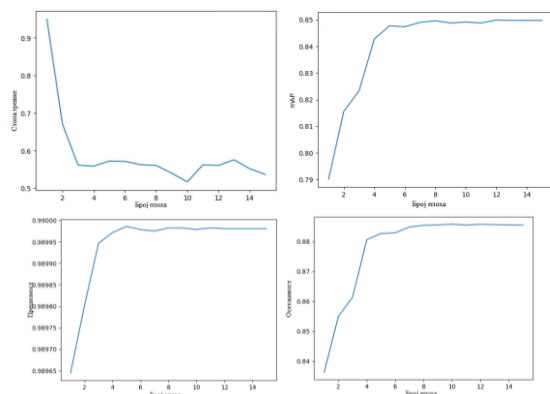
Slika 4. Segmentacija nakon doobučavanja nad skupom od 100 snimaka (desno, nakon NMS)



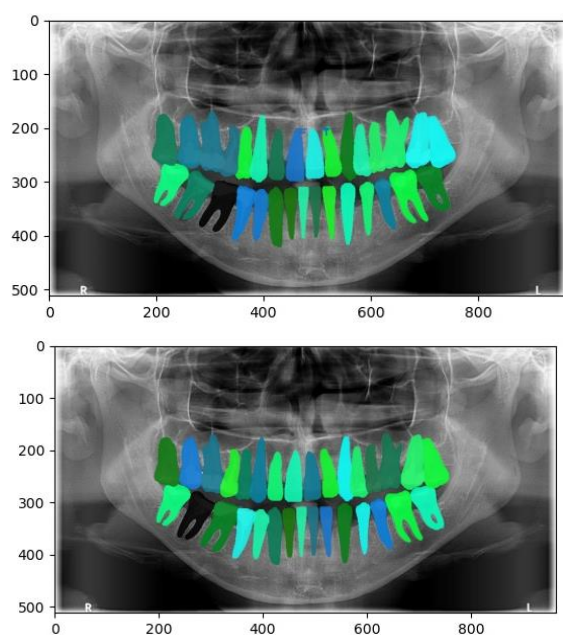
Slika 5. Karakteristike detektora nakon doobučavanja nad skupom od 100 snimaka



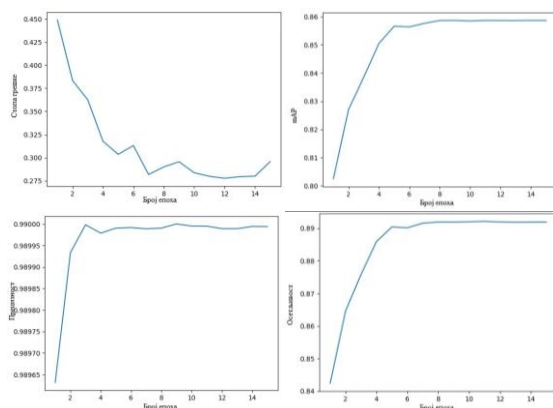
Slika 6. Segmentacija nakon doobučavanja nad skupom od 1000 snimaka (desno, nakon NMS)



Slika 7. Karakteristike detektora nakon doobučavanja nad skupom od 1000 snimaka



Slika 8. Segmentacija nakon doobučavanja nad skupom od 2000 snimaka (dole, nakon NMS)



Slika 9. Karakteristike detektora nakon doobučavanja nad skupom od 2000 snimaka

4. ZAKLJUČAK

Cilj ovog rada bio je razvoj metode za pouzdanu detekciju i segmentaciju zuba u panoramskim radiografskim snimcima vilice [6], uz što manje uloženo vreme i resurse za obuku modela. Pokazano je da se u slučaju ograničenih resursa model može uspešno konstruisati i korišćenjem manjeg skupa podataka, uz relativno zadovoljavajuće rezultate. Polazna pretpostavka da je u svrhu rešavanja navedenog zadatka moguće koristiti predloženi pristup zasnovan na doobučavanju postojećeg modela, prethodno obučavanog na slikama i objektima drugačijeg modaliteta, uspešno je testirana i potvrdila je da su diskriminativna obeležja na kojima se zasniva prepoznavanje opšteg karaktera.

ZAHVALNICA

Zahvaljujemo se doktoru Bojanu Petroviću, redovnom profesoru Medicinskog fakulteta Univerziteta u Novom Sadu, i Klinici za stomatologiju Vojvodine, koji su kroz konsultacije i pripremu panoramskih radiografskih snimaka podržali istraživanje opisano u radu. Takođe se zahvaljujemo i kolegama sa Federalnog univerziteta u Bahiji, u Brazilu, koji su ljubazno ustupili bazu slika [4] za potrebe ovog istraživanja.

5. LITERATURA

- [1] Y. LeCun, Y. Bengio, G. Hinton (2015). Deep learning. Nature, 28;521(7553):436-44. doi: 10.1038/nature14539. PMID: 26017442.
- [2] Lin, T. et al. (2014). Microsoft COCO: Common Objects in Context. ECCV 2014. Lecture Notes in Computer Science, vol 8693. Springer
- [3] A. Dutta and A. Zisserman. (2019). The VIA Annotation Software for Images, Audio and Video. In Proceedings of the 27th ACM International Conference on Multimedia (MM '19), October 21–25, Nice, France. ACM, New York, NY, USA, <https://doi.org/10.1145/3343031.3350535>.
- [4] B. Silva, L. Pinheiro, B. Sobrinho, F. Lima, B. Sobrinho, K. Abdalla, M. Pithon, P. Cury, L. Oliveira, "OdontoAI: A human-in-the-loop labeled data set and an online platform to boost research on dental panoramic radiographs", arXiv.org, 2022.[Online]. <https://arxiv.org/abs/2203.15856>
- [5] K. He, G. Gkioxari, P. Dollár and R. Girshick, (2017). "Mask R-CNN," IEEE International Conference on Computer Vision, Venice, Italy, pp. 2980-2988, doi: 10.1109/ICCV.2017.322.
- [6] G. Jader, J. Fontineli, M. Ruiz, K. Abdalla, M. Pithon and L. Oliveira, (2018). "Deep Instance Segmentation of Teeth in Panoramic X-Ray Images," 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), Parana, Brazil, 2018, pp. 400-407, doi: 10.1109/SIBGRAPI.2018.00058.

Kratka biografija:



Dorina Petrović rođena je u Zrenjaninu 1994. god. Osnovne studije Energetike, elektronike i telekomunikacija na Fakultetu tehničkih nauka u Novom Sadu završila je 2017. godine. Master rad na Fakultetu tehničkih nauka iz oblasti Elektrotehnike i računarstva – Obrada signala, uspešno je odbranila 2023.

email adresa : dorina.petrovic@me.com