

Zbornik strokovne konference

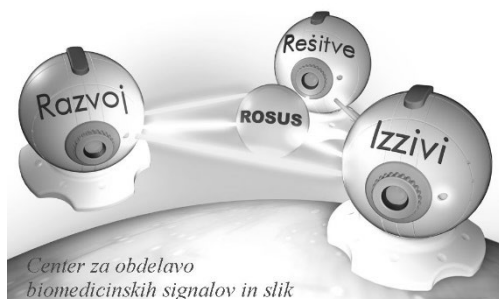
ROSUS 2025

Računalniška obdelava slik in
njena uporaba v Sloveniji 2025

Urednik:
Božidar Potočnik



Maribor, 20. marec 2025



ROSUS 2025 – Računalniška obdelava slik in njena uporaba v Sloveniji 2025

Zbornik 19. strokovne konference

Urednik
Božidar Potočnik

Marec 2025

Naslov <i>Title</i>	ROSUS 2025 - Računalniška obdelava slik in njena uporaba v Sloveniji 2025 ROSUS 2025 - <i>Computer image processing and its application in Slovenia 2025</i>
Podnaslov <i>Subtitle</i>	Zbornik 19. strokovne konference <i>Proceedings of the 19th Professional Conference</i>
Urednik <i>Editor</i>	Božidar Potočnik (Univerza v Mariboru, Fakulteta za elektrotehniko, računalništvo in informatiko)
Recenzija <i>Review</i>	Aleš Holobar (Univerza v Mariboru, Fakulteta za elektrotehniko, računalništvo in informatiko) Simon Dobrišek (Univerza v Ljubljani, Fakulteta za elektrotehniko)
Tehnični urednik <i>Technical editor</i>	Jan Perša (Univerza v Mariboru, Univerzitetna založba)
Oblikovanje ovitka <i>Cover designer</i>	Gašper Sedej (Univerza v Mariboru, Fakulteta za elektrotehniko, računalništvo in informatiko)
Grafične priloge <i>Graphic material</i>	Viri so lastni, razen če ni navedeno drugače. Avtorji prispevkov in Potočnik (urednik), 2025
Konferenca <i>Conference</i>	Strokovna konferenca ROSUS 2025: Računalniška obdelava slik in njena uporaba v Sloveniji 2025
Datum in kraj konference <i>Conference date & location</i>	20. marec 2025, Maribor, Slovenija
Programski odbor <i>Program committee</i>	Aleš Holobar, Damjan Zazula, Boris Cigale, Marjan Mernik, Peter Peer, Tomaž Tomažič, Martin Šavc, Slobodan Ribarić, Vojko Flis, Saša Divjak, Stanislav Kovačič, Zdravko Kačič, Aleš Leonardiš, Simon Dobrišek, Vitomir Štruc, Franc Solina, Andrej Šoštarič, Franjo Pernuš, Jure Skvarč, Božidar Potočnik
Organizacijski odbor <i>Organization committee</i>	Gašper Sedej, Jurij Munda, Uroš Mlakar, Matjaž Divjak, Martin Šavc, Danilo Korže, Nina Murks, Matej Kramberger, Leon Kutoš, Aleš Holobar, Božidar Potočnik
Založnik <i>Published by</i>	Univerza v Mariboru Univerzitetna založba Slomškov trg 15, 2000 Maribor, Slovenija https://press.um.si , zalozba@um.si
Izdajatelj <i>Issued by</i>	Univerza v Mariboru Fakulteta za elektrotehniko, računalništvo in informatiko Koroška cesta 46, 2000 Maribor, Slovenija https://feri.um.si , feri@um.si

Izdaja
Edition Prva izdaja

Vrsta publikacije
Publication type E-knjiga

Dostopno na
Available at <https://press.um.si/index.php/ump/catalog/book/957>

Izdano
Published Maribor, marec 2025



© **Univerza v Mariboru, Univerzitetna založba**
University of Maribor, University Press

Besedilo/ Text
© Avtorji in Potočnik, 2025

To delo je objavljeno pod licenco Creative Commons Priznanje avtorstva-Deljenje pod enakimi pogoji 4.0 Mednarodna. / *This work is licensed under the Creative Commons Attribution-ShareAlike 4.0 International License.*

Uporabnikom se dovoli reproduciranje, distribuiranje, dajanje v najem, javno priobčitev in predelavo avtorskega dela, če navedejo avtorja in širijo avtorsko delo/predelavo naprej pod istimi pogoji. Za nova dela, ki bodo nastala s predelavo, je tudi dovoljena komercialna uporaba.

Vsa gradiva tretjih oseb v tej knjigi so objavljena pod licenco Creative Commons, razen če to ni navedeno drugače. Če želite ponovno uporabiti gradivo tretjih oseb, ki ni zajeto v licenci Creative Commons, boste morali pridobiti dovoljenje neposredno od imetnika avtorskih pravic.

<https://creativecommons.org/licenses/by-sa/4.0/>

CIP - Kataložni zapis o publikaciji
Univerzitetna knjižnica Maribor
004.932(497.4)(082)(0.034.2)

STROKOVNA konferenca ROSUS (konferenca) (19 ; 2025 ; Maribor)
ROSUS 2025 [Elektronski vir] : računalniška obdelava slik in njena uporaba v Sloveniji 2025 : zbornik 19. strokovne konference : [20. marec 2025, Maribor, Slovenija] / urednik Božidar Potočnik. - 1. izd. - E-zbornik. - Maribor : Univerza v Mariboru, Univerzitetna založba, 2025

Način dostopa (URL): <https://press.um.si/index.php/ump/catalog/book/957>
ISBN 978-961-286-960-1 (PDF)
doi: 10.18690/um.feri.2.2025
COBISS.SI-ID 227936771

ISBN 978-961-286-960-1 (pdf)
978-961-286-961-8 (mehka vezava)

DOI <https://doi.org/10.18690/um.feri.2.2025>

Cena
Price Brezplačni izvod

Odgovorna oseba založnika
For publisher prof. dr. Zdravko Kačič,
rektor Univerze v Mariboru

Citiranje
Attribution Potočnik, B. (ur.). (2025). *ROSUS 2025 – Računalniška obdelava slik in njena uporaba v Sloveniji 2025: zbornik 19. strokovne konference*. Univerza v Mariboru, Univerzitetna založba. doi: 10.18690/um.feri.2.2025

POKROVITELJI



Slovenska sekcija



ROSUS 2025

<http://rosus.feri.um.si>

Kazalo

	Predgovor <i>Foreword</i> Božidar Potočnik	1
VABLJENA PREDAVANJA		3
	Profesor in podjetnik: Primer podjetja Sensum <i>Professor and Entrepreneur: The Sensum Case</i> Franjo Pernuš	5
	Razvoj simulacijskega pogona <i>Simulation Engine Development</i> Vito Čuček	7
	Percepcija pri avtonomni vožnji za »off-road« vozila <i>Perception in Autonomous Driving for Off-Road Vehicles</i> Jan Ogrin	11
DOPOLDANSKA IN POPOLDANSKA SEKCIJA Industrijske aplikacije, medicinske in biomedicinske aplikacije, študentske aplikacije, drugo		17
1	Avtomatiziran postopek kalibracij kamer satelita TRISAT-S <i>Automated Calibration Process for TRISAT-S Satellite Cameras</i> Emanuel Nikl Hutinski	19
2	Exploring the Differences in Pruning Methods for Convolutional Neural Networks <i>Raziskovanje razlik med metodami obrezovanja za konvolucijske nevronske mreže</i> Romanela Lajić, Peter Peer, Žiga Emeršič	31
3	Generiranje sintetičnih CT slik iz MR slik področja glave in vratu z uporabo difuzijskih modelov <i>Generation of Synthetic CT Images From MR Images in the Head and Neck Region Using Diffusion Models</i> Rok Marko Šter, Gašper Podobnik, Tomaž Vrtovec	41

4	AIM@VET-Inspired University Level Education Strategies for Teaching Comp-Uter Vision and Biometrics <i>AIM@VET izobraževalni materiali za univerzitetno poučevanje računalniškega vida in biometrije</i> Žiga Emeršič, Gregor Hrastnik, Nataša Meh Peer, Peter Peer	55
5	Segmentacija regij zanimanja v histopatoloških posnetkih kolorektalnih polipov <i>Region of Interest Segmentation in Histopathological Images of Colorectal Polyp</i> Martin Šavc, Božidar Potočnik	65
6	Uporaba metod računalniškega vida za pomoč pri telesni vadbi <i>Use of Computer Vision Methods for Guidance in Physical Exercises</i> Blaž Černi, Borut Batagelj	77
7	Unexploded Ordnance Detection in Hyperspectral Images by Using Deep Neural Networks <i>Detektiranje neeksplodiranih ubojnih sredstev na hiperspektralnih slikah z uporabo globokih nevronske mreže</i> Milan Bajić, Božidar Potočnik	93
8	Analiza odbojarske igre z uporabo algoritmov računalniškega vida <i>Volleyball Game Analysis Using Computer Vision Algorithms</i> Marko Plankelj, Uroš Mlakar	107

Spoštovani!

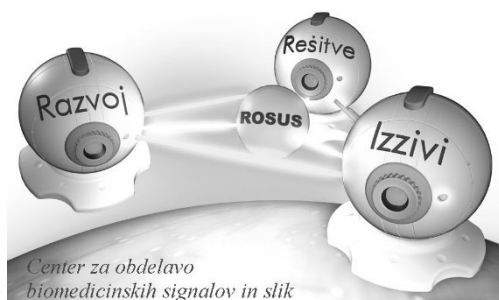
BOŽIDAR POTOČNIK

predsednik konference ROSUS 2025

Po osemnajstih konferencah ROSUS 2006–2024 s konferenco ROSUS 2025 nadaljujemo promoviranje ekonomske koristi računalniške obdelave slik in računalniškega vida na področjih industrije, biomedicine in drugih poslovnih procesov. Vezi, ki smo jih na prejšnjih konferencah stkali med raziskovalci, razvijalci, ponudniki rešitev ter uporabniki, želimo dodatno okrepiti, ob tem pa nuditi možnosti sklepanja novih sodelovanj in svetovanja pri razreševanju konkretnih poslovnih oziroma raziskovalnih problemov.

Glavni poudarek namenjamo aplikacijam s področja računalniške obdelave slik, ki so že integrirane oziroma pripravljene za integracijo v poslovne procese. Demonstrirali bomo, da avtomatska obdelava v industriji lahko zaradi svoje natančnosti in hitrosti prinaša velike ekonomske koristi, hkrati pa nakazali, da takšne aplikacije nudijo nove priložnosti za uveljavitev na trgu visokih tehnologij. Ne smemo pozabiti še na možnost diskusije ter predstavitev konkretnih problemov in potreb, ki se porajajo pri uporabnikih, s katerimi bomo računalniško obdelavo slik in njeno koristnost še bolj približali avditoriju.

Naj sklenemo uvodne misli s prisrčno zahvalo Javnemu skladu Republike Slovenije za podjetništvo, ki v okviru konference ROSUS 2025 predstavlja zanimive finančne instrumente za spodbujanje prenosa tehnoloških rešitev v podjetniško sfero. Iskreno se zahvaljujemo tudi podjetju Audax za sprejem pokroviteljstva konference. Izpostaviti želimo še medijske pokrovitelje revijo IRT3000, revijo Svet elektronike, revijo Ventil in revijo Avtomatika + elektronika, ki so intenzivno promovirali konferenco ROSUS 2025 ter pomen strojnega vida v slovenskem prostoru.



VABLJENA PREDAVANJA

PROFESOR IN PODJETNIK: PRIMER PODJETJA SENSUM

FRANJO PERNUŠ

Univerza v Ljubljani, Fakulteta za elektrotehniko, Ljubljana, Slovenija
franjo.pernus@fe.uni-lj.si

V zadnjem času so se številne univerze preoblikovale iz tradicionalne izobraževalne in raziskovalne univerze v podjetniško univerzo, ki spodbuja podjetniške dejavnosti zaposlenih in študentov. Vloga podjetniške univerze torej ni le izobraževanje in ustvarjanje novega znanja, temveč tudi podpiranje ustanavljanja na novem znanju temelječih podjetij, od katerih bi imela korist celotna družba. Vprašanje pa je, ali so univerzitetni učitelji in sodelavci po naravi tudi podjetniki. V prispevku podajamo nekaj podobnosti med delom, ki ga opravljajo učitelji in podjetniki, in s tem skušamo odgovoriti na zgornje vprašanje. V nadaljevanju opisujemo primer podjetja Sensum, ki so ga leta 2000 ustanovili učitelj in trije študenti. Podjetje Sensum je specializirano za razvoj, proizvodnjo in trženje avtomatskih sistemov za vizualno pregledovanje v farmacevtski industriji in industriji prehranskih dopolnil. Sensum ponuja rešitve visoke in srednje zmogljivosti za 100 % pregledovanje kakovosti tablet, neprozornih, prozornih in mehko želatinastih kapsul ter procesno analitične tehnologije za vizualno spremljanje farmacevtskih procesov v realnem času.

ISBN
978-961-286-960-1

Ključne besede:
strojni vid,
biomedicina,
industrijske aplikacije,
kontrola kvalitete,
prenos znanja



Univerzitetna založba
Univerze v Mariboru

ISBN
978-961-286-960-1

PROFESSOR AND ENTREPRENEUR: THE SENSUM CASE

FRANJO PERNUŠ

University of Ljubljana, Faculty of Electrical Engineering, Ljubljana, Slovenia
franjo.pernus@fe.uni-lj.si

Keywords:
Machine Vision,
Biomedicine,
industrial applications,
quality control,
knowledge transfer

Recently, many universities have transformed themselves from a traditional teaching and research university to an entrepreneurial university that encourages entrepreneurial activities of staff and students. The role of an entrepreneurial university is therefore not only to educate and create new knowledge, but also to support the creation of new knowledge-based enterprises that would benefit society as a whole. The question is whether university teachers and colleagues are entrepreneurs by nature. In this article, we present some similarities between the work of teachers and that of entrepreneurs in an attempt to answer the above question. Below we describe the case of Sensum, a company founded in 2000 by a teacher and three students. Sensum specialises in the development, production and marketing of automatic visual inspection systems for the pharmaceutical and food supplement industries. Sensum offers high and medium capacity solutions for 100% quality inspection of tablets, opaque, transparent and soft gelatine capsules, as well as process analytical technologies for real-time visual monitoring of pharmaceutical processes.



University of Maribor Press

RAZVOJ SIMULACIJSKEGA POGONA

VITO ČUČEK

Guardiaris, Ljubljana, Slovenija
cucekv@guardiaris.com

Podjetje Guardiaris se ukvarja z razvojem vojaških trenažerjev in civilnih simulatorjev. Gonilno logiko proizvedenih fizičnih trenažerjev in naprav poganja v lastni režiji razvit specializiran simulacijski in grafični pogon. Razvit je bil z namenom lažje integracije in optimizacije z namensko strojno opremo ter posebnosti, ki tovrstne aplikacije razlikujejo od komercialnih igralnih pogonov. Predstavili vam bomo nekaj tehničnih izzivov in rešitev s katerimi se soočamo. Obravnavali bomo pristope pri programskem projektiranju trenažerjev, simulacijo fizike projektilov in penetracij, procese proceduralnega generiranja terena na osnovi geodetskih podatkov ter rešitve za vizualizacijo obsežnih terenov. Obrazložili bomo posebnosti simulacijskih pogonov, ki za razliko od igralnih zahtevajo širšo povezljivost z ostalimi sistemi ter potrebe integrirane možnosti časovne analize poteka.

ISBN
978-961-286-960-1

Ključne besede:
simulacijski pogon,
grafični pogon,
vojaški trenažer,
napredna vizualizacija,
razvoj programske opreme



Univerza Tehna založba
Univerze v Mariboru

ISBN
978-961-286-960-1

Keywords:
simulation engine,
graphics engine,
military simulator,
advanced visualization,
software development

SIMULATION ENGINE DEVELOPMENT

VITO ČUČEK

Guardiaris, Ljubljana, Slovenia
cucekv@guardiaris.com

Guardiaris is engaged in the development of military and civil simulators. The control logic of the physical simulators and devices produced is driven by a specialised simulation and graphics engine developed in-house. It has been developed with the aim of easier integration and optimisation with dedicated hardware and the specific features that distinguish such applications from commercial game engines. We will present some of the technical challenges and solutions we have faced. We will discuss approaches to software design of simulators, simulation of projectile physics and penetration, procedural terrain generation methods based on geodetic data, and solutions for visualisation of large terrains. We will explain the specific characteristics of simulation engines, which, unlike game engines, require wider connectivity with other systems and the need for integrated time-series analysis capabilities.



University of Maribor Press

1 Uvod

V tem predavanju bo predstavljena izgradnja realnočasovnega sintetičnega terena s tehniko združevanja letalskih posnetkov in vektorskih podatkov za izgradnjo proceduralnih tekstur in objektov. Dotaknili se bomo dveh pristopov izrisovanja oddaljenega območja ter tehnike virtualnih tekstur za potrebe pretočnega posodabljanje in izrisa obsežnih terenov.

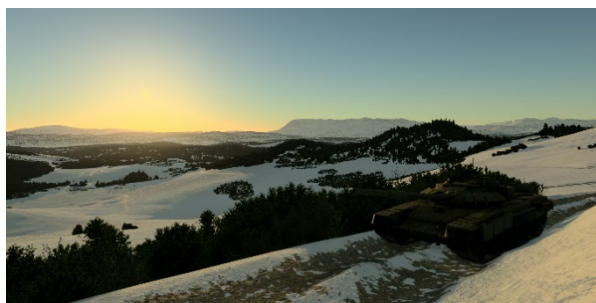
Na sliki 1 in sliki 3 vidimo prehod med proceduralno generiranim terenom na podlagi vektorskih podatkov v oddaljen teren. Vse prikazane slike so ustvarjene z direktnim izvozom ob realnočasovnem predvajanju simulacije.



Slika 1: Proceduralno generiran teren

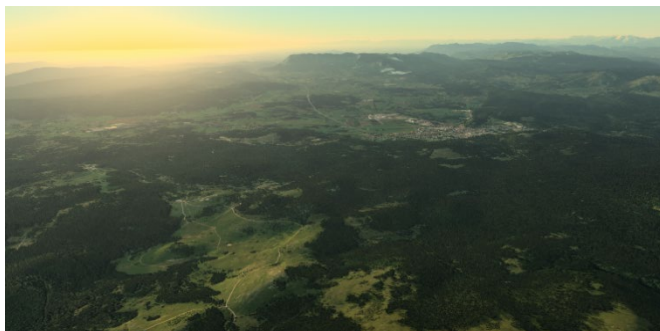
Vir: lasten

Na sliki 2 je prikazan učinek snega, kot primer proceduralnega generiranja z uporabo virtualnih tekstur.



Slika 2: Dodan učinek snega v sceno

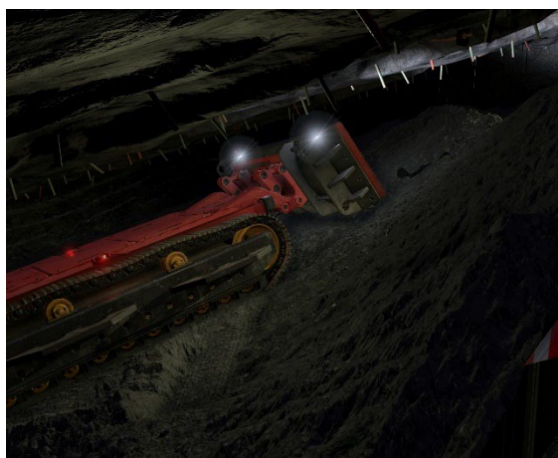
Vir: lasten



Slika 3: Slika oddaljenega terena

Vir: lasten

Na sliki 4 pa vidimo prikaz simulacije sipanja in odmetavanja grušča, namenjeni za izdelavo trenažerja daljinsko vodenega robota za odmetavanje materiala v rudniku.



Slika 4: Simulacija sipanja in odmetavanja grušča

Vir: lasten

Na predavanju bomo obrazložili tudi ostale optimizacijske metode algoritmov, potrebne za simulacijo in prikaz virtualnih okolij, ter pristopov izdelave trenažerjev.

PERCEPCIJA PRI AVTONOMNI VOŽNJI ZA »OFF-ROAD« VOZILA

JAN OGRIN

Guardiaris, Ljubljana, Slovenija
ogrinj@guardiaris.com

Predavanje sega na področje tehnologije zaznavanja in planiranja poti za namen avtonomne vožnje modularnega oklepnega vozila, kar omogoča varno in učinkovito navigacijo v kompleksnih, neoznačenih in nepredvidljivih okoljih. Senzorji, ki se uporabljajo v tem procesu, so barvne kamere, stereo kamere, termalne kamere, lidarji, radarji, inercialni senzorji in GPS. Poseben poudarek predavanja je na izzivih kot so kalibracija in sinhronizacija senzorjev ter seveda uporaba podatkov iz naštetih senzorjev za razvoj in testiranje algoritmov z namenom hkratne lokalizacije in mapiranja okolja.

ISBN
978-961-286-960-1

Ključne besede:
avtonomna vožnja,
senzorski podatki,
planiranje poti,
zaznavanje terena,
kalibracija podatkov



Univerzitetna založba
Univerze v Mariboru

ISBN
978-961-286-960-1

PERCEPTION IN AUTONOMOUS DRIVING FOR OFF-ROAD VEHICLES

JAN OGRIN

Guardiaris, Ljubljana, Slovenia
ogrinj@guardiaris.com

Keywords:
autonomous driving,
sensor data,
route planning,
terrain detection,
data calibration

The lecture covers the field of sensor and path planning technology for the purpose of autonomous driving of a modular armoured vehicle, enabling safe and efficient navigation in complex, unmarked and unpredictable environments. Sensors used include colour cameras, stereo cameras, thermal cameras, lidars, radars, inertial sensors and GPS. The talk will focus on challenges such as sensor calibration and synchronisation and, of course, the use of data from the listed sensors to develop and test algorithms for simultaneous localisation and mapping of the environment.



University of Maribor Press

1 Uvod

Eden od raziskovalnih projektov, s katerim se ukvarjamo pri podjetju Guardiaris, je večnamensko 6x6 oklepno vozilo s hibridnim dizelsko-električnim pogonom brez posadke. Oklepno vozilo je možno upravljati na daljavo preko radijske povezave, lahko pa se vozi tudi v avtonomnem načinu. Omogoča nosilnost do 3000 kg. Slednje pomeni, da ga lahko opremimo s strelnim orožjem, opremo za gašenje ali pa se uporablja za prevoz tovora. Električni motorji omogočajo tiho vožnjo ob premagovanju zahtevnih terenov.

Slike od 1 do 4 predstavljajo idejno zasnovo vozila in demonstracijo njegove uporabe in/ali upravljanja.



Slika 1: Idejna zasnova oklepnega vozila

Vir: lasten



Slika 2: Model oklepnega vozila

Vir: lasten



Slika 3: Model oklepnega vozila v simuliranem okolju

Vir: lasten



Slika 4: Demonstracija uporabe in/ali upravljanja oklepnega vozila na sejmskem dogodku

Vir: lasten

V Guardiarisu razvijamo postopke za kalibracijo in sinhronizacijo senzorjev, kot so kamere, ki zajemajo sceno v vidnem in termalnem spektru, pasivne stereo kamere, lidarji, radarji, žiroskopi in pospeškometri ter naprave GPS. Podatki iz teh senzorjev vozilu omogočajo percepcijo in mapiranje okolja, posledično pa varno in učinkovito planiranje poti. Sočasno poteka razvoj digitalnega dvojčka, trenažerja in simulatorja, za namen lažjega treniranja upravljalcev vozila.

Na sliki 5 je prikazano prototipno vozilo, ki se uporablja za razvoj algoritmov kalibracije, percepcije, lokalizacije in planiranja poti.



Slika 5: Prototipno vozilo za razvoj algoritmov

Vir: lasten

DOPOLDANSKA IN POPOLDANSKA SEKCIJA

Industrijske aplikacije
Medicinske in biomedicinske aplikacije
Študentske aplikacije
Drugo

AVTOMATIZIRAN POSTOPEK KALIBRACIJ KAMER SATELITA TRISAT-S

EMANUEL NIKL HUTINSKI

Univerza v Mariboru, Fakulteta za elektrotehniko, računalništvo in informatiko, Maribor,
Slovenija
emanuel.hutinski@um.si

V prispevku predstavljamo metodologijo avtomatizirane kalibracije kamer, ki so nameščene na satelitu TRISAT-S. Sistem temelji na predhodnih izkušnjah z misijo TRISAT-R, kjer so bile miniaturne kamere uporabljene za tehnološko demonstracijo. Za zagotavljanje natančne rekonstrukcije slikovnih podatkov je bila razvita popolnoma avtomatizirana metoda kalibracije, ki vključuje robotsko roko za precizno nastavitvev položajev kamer in optimizacijske algoritme za korekcijo optičnih distorzij. Pri kalibraciji se uporabljajo Förstner-jev operater za detekcijo kalibracijskih vzorcev, Levenberg-Marquardt metoda za minimizacijo reprojekcijske napake in geometrijski model kamere na podlagi »pinhole« projekcije. Poleg kalibracije predstavljamo tudi koncept 360-stopinjske rekonstrukcije okolja, ki temelji na združevanju slik iz vseh šestih kamer, nameščenih na satelitu. Pri tem je ključnega pomena geometrijska poravnava slik, korekcija distorzije in združevanje podatkov v enotno panoramsko predstavitev. Rezultati kažejo, da je predlagan sistem robusten in ponovljiv, s čimer omogoča visoko natančnost in avtomatizirano obdelavo podatkov brez človeškega posredovanja. Nadaljnje raziskave bodo usmerjene v implementacijo metode na dejanskem satelitu in analizo učinkovitosti rekonstrukcije v realnih vesoljskih razmerah.

DOI
[https://doi.org/
10.18690/um.feri.2.2025.1](https://doi.org/10.18690/um.feri.2.2025.1)

ISBN
978-961-286-960-1

Ključne besede:
kalibracija kamere,
avtomatizirana kalibracija,
geometrijska poravnava
slike,
Förstnerjev operater,
optimizacija Levenberg-
Marquardt,
model »pinhole« kamere,
korekcija optičnega
popačenja



Univerzitetna založba
Univerze v Mariboru

DOI
[https://doi.org/
10.18690/um.feri.2.2025.2](https://doi.org/10.18690/um.feri.2.2025.2)

ISBN
978-961-286-960-1

Keywords:

camera calibration,
automated calibration,
satellite imaging,
geometric image alignment,
Förstner operator,
Levenberg-Marquardt
optimization,
Pinhole camera model,
optical distortion correction

AUTOMATED CALIBRATION PROCESS FOR TRISAT-S SATELLITE CAMERAS

EMANUEL NIKL HUTINSKI

University of Maribor, Faculty of Electrical Engineering and Computer Science, Institute
of Computer Science, Maribor, Slovenia
emanuel.hutinski@um.si

In this paper, we present the methodology for the automated calibration of cameras mounted on the Trisat-S satellite. The system builds upon previous experience from the TRISAT-R mission, where miniature cameras were used for a technological demonstration. To ensure accurate reconstruction of image data, a fully automated calibration method was developed, incorporating a robotic arm for precise camera positioning and optimization algorithms for optical distortion correction. The calibration process employs the Förstner operator for calibration pattern detection, the Levenberg-Marquardt method for minimizing reprojection error, and the geometric pinhole camera model. In addition to calibration, we introduce the concept of 360-degree environmental reconstruction, which is based on merging images from all six cameras mounted on the satellite. Key aspects of this process include geometric alignment of images, distortion correction, and data fusion into a unified panoramic representation. The results demonstrate that the proposed system is robust and repeatable, enabling high accuracy and fully automated data processing without human intervention. Future research will focus on implementing the method on the actual satellite and analyzing the effectiveness of the reconstruction in real space conditions.



1 Uvod

Na misiji TRISAT-R, ki je potekala v osrčju ionosfere in notranjega Van Allenovega sevalnega pasu, sta se testirali dve miniaturni kameri, ki sta služili kot tehnološka demonstracija uporabe visoko miniaturne tehnologije slikanja. Kameri sta bili del eksperimenta zajemanja »učinka črnega sonca«, dodatno pa sta posneli sliko planeta Zemlja z izjemno nizko ločljivostjo 320x320 slikovnih točk. Misija TRISAT-S bo uporabila enake kamere, ki bodo nameščene na vse stranice satelita, tako da bomo zajeli slike iz vseh zornih kotov. Postopek rekonstrukcije poteka v naslednjih korakih:

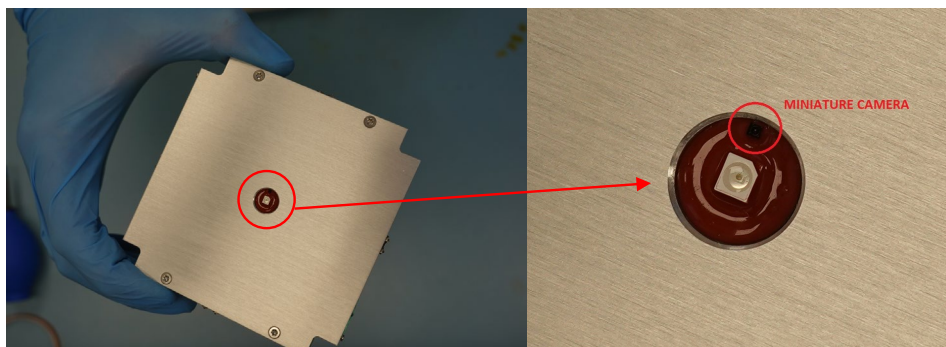
1. Zajem slike
2. Kalibracija sistema
3. Združevanje podatkov
4. Rekonstrukcija

V tem članku smo se osredotočili na kalibracijo sistema, ki je ključen postopek pri zagotavljanju natančnosti in zanesljivosti slikovnih podatkov v različnih aplikacijah, od industrijskih in znanstvenih sistemov do navigacije, daljinskega zaznavanja in računalniškega vida. Gre za sistematičen proces določanja in popravljanja optičnih in geometrijskih napak, ki nastanejo zaradi fizičnih lastnosti optičnih komponent in delovnih pogojev kamere. Geometrijska kalibracija identificira odnos med 3D svetom in 2D slikovnimi podatki, pri čemer se odpravljajo sistematične napake v preslikavi prostora. Ključni cilji geometrijske kalibracije so:

1. Določitev notranjih parametrov kamere (intrinzična kalibracija)
2. Določitev zunanjih parametrov (ekstrinzična kalibracija)

Ti parametri so bistveni pri aplikacijah, kot so 3D rekonstrukcija, robotski vid, navigacija avtonomnih sistemov in fotogrametrija.

Za doseganje najboljših rezultatov mora biti kalibracija izvedena v enakih pogojih za vse kamere, kar vključuje kote zajemanja, osvetlitev in druge vplivne dejavnike. Zaradi tega smo razvili popolnoma avtomatiziran sistem kalibracije kamer z uporabo robotske roke, ki omogoča natančno in ponovljivo nastavitve vsake kamere v sistemu.



Slika 1: Miniaturna kamere satelita TRISAT-R.

Vir: lasten.

2 Matematični model kamere

Standardni matematični model perspektivne projekcije temelji na »pinhole« modelu kamere (Sturm, 2014), ki opisuje transformacijo 3D točk v 2D slikovne koordinate. Ta model je osnovna oblika kamere brez leč, kjer svetloba prehaja skozi majhno luknjo in ustvari obrnjeno sliko na nasprotni strani. »Pinhole« model lahko opišemo z naslednjimi matematičnimi enačbami:

$$\text{Projekcijske enačbe: } x = f \cdot \frac{X}{Z} ; y = f \cdot \frac{Y}{Z}$$

$$\text{Homogene koordinate: } \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \sim \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

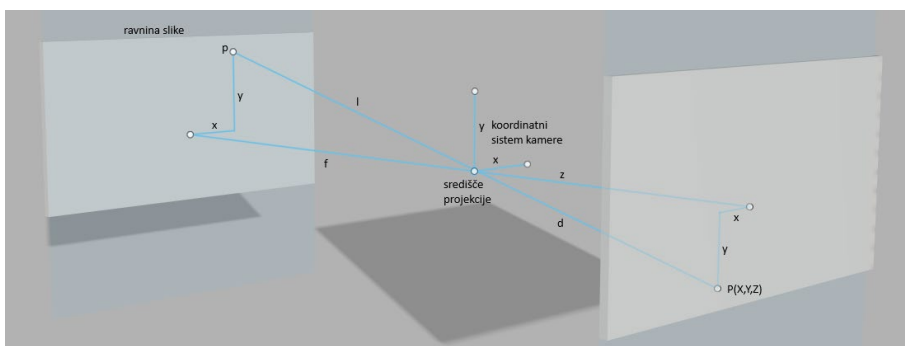
$$\text{Intrizični parametri: } K = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

$$\text{Ekstrinzični parametri: } \begin{bmatrix} R_{3 \times 3} & T_{3 \times 1} \\ 0_{1 \times 3} & 1 \end{bmatrix}$$

Kjer je f goriščna razdalja kamere, f_x in f_y pa goriščni razdalji v slikovnih enotah v smeri x in y . c_x in c_y sta koordinati središča slike v slikovnih koordinatah. Matrika R predstavlja rotacijsko transformacijo, translacijski vektor T pa premik med koordinatnim sistemom sveta in koordinatnim sistemom kamere. Skupaj opisujeta transformacijo iz 3D koordinat sveta v 3D koordinate kamere. V resničnih kamerah

so parametri premika slikovnega središča in razmerje stranic slikovnega elementa. Ti parametri se upoštevajo v intrinzični matriki kamere, zato lahko zapišemo enačbo:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = K \cdot [R \quad T] \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$



Slika 2: Preslikava slike z modelom »Pinhole« kamere

Vir: lasten

3 Korekcija distorzije

Realne kamere ne sledijo idealnemu »pinhole« modelu, saj optični sistemi povzročajo nelinearne popačitve slike. Najpogosteje se modelirata radialna in tangencialna distorzija.

Radialna distorzija povzroča ukrivljanje ravnih črt, ki je najbolj izrazito proti robovom slike. Popravlja se z modelom:

$$x_{popaceno} = x_{idealno} (1 + k_1 r^2 + k_2 r^4 + k_3 r^6)$$

$$y_{popaceno} = y_{idealno} (1 + k_1 r^2 + k_2 r^4 + k_3 r^6)$$

Tangencialna distorzija nastane zaradi nepopolne poravnave leč in tipala ter se modelira z izrazi:

$$x_{popaceno} = x_{idealno} \cdot 2p_1 xy + p_2 (r^2 + 2x^2)$$

$$y_{popaceno} = y_{idealno} \cdot p_1 (r^2 + 2y^2) + 2p_2 xy$$

Pri čemer je » r « evklidska razdalja med točko popačene slike in središčem popačenja:

$$r^2 = x_{idealno}^2 + y_{idealno}^2$$

Koeficienti k_1 , k_2 in k_3 so koeficienti radialnega popačenja, koeficienta p_1 in p_2 pa koeficienta tangencialnega popačenja.

4 Algoritem kalibracije kamere

Za izračun kalibracijskih parametrov se uporabljajo različni numerični algoritmi, med katerimi sta najpogostejši metoda Zhang in metoda na podlagi ArUco markerjev.

Metoda Zhang uporablja šahovnico kot kalibracijsko tarčo. Metoda vključuje:

1. Detekcijo kotov na šahovnici (metoda s Forsterjevim operatorjem)
2. Izračun homografije med kalibracijsko tarčo in slikovno ravnino
3. Reševanje sistema enačb za določitev intrinzičnih in ekstrinzičnih parametrov
4. Nelinearno optimizacijo (npr. Levenberg-Marquardt) za zmanjšanje napake reprojekcije

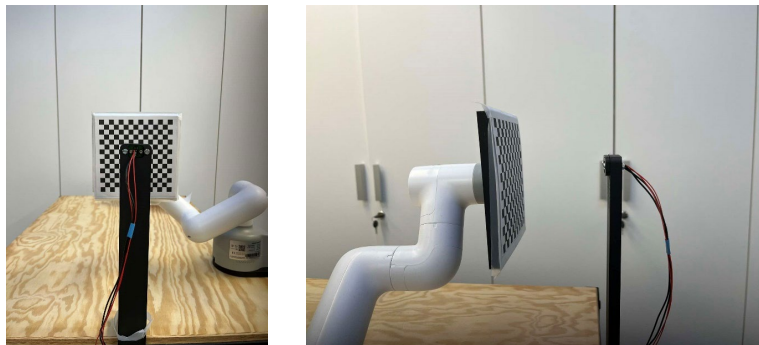
Alternativno je mogoče uporabiti ArUco markerje, ki zagotavljajo robustnejše zaznavanje v zahtevnih svetlobnih pogojih in omogočajo samodejno identifikacijo referenčnih točk. Vendar ta metoda ni bila primerna za obravnavani sistem, saj nizka ločljivost kamere ni omogočala zadostne natančnosti pri detekciji ArUco markerjev na kalibracijski tarči.

5 Metodologija avtomatizirane kalibracije

Avtomatizirana kalibracija kamer vključuje integracijo algoritmov v sistem, ki zmanjšuje potrebo po ročnem posredovanju. Ključni koraki vključujejo:

1. Samodejno spreminjanje položaja kalibracijske tarče

Uporablja se robotska roka MyCobot 280 (Oosterwyck, 2018). Ponovljivost gibanja je zagotovljena s CSV datoteko, v kateri so definirani položaji vseh sklepov za 27 različnih pozicij. Gibanje je povsem avtomatizirano.



Slika 3: Realni sistem

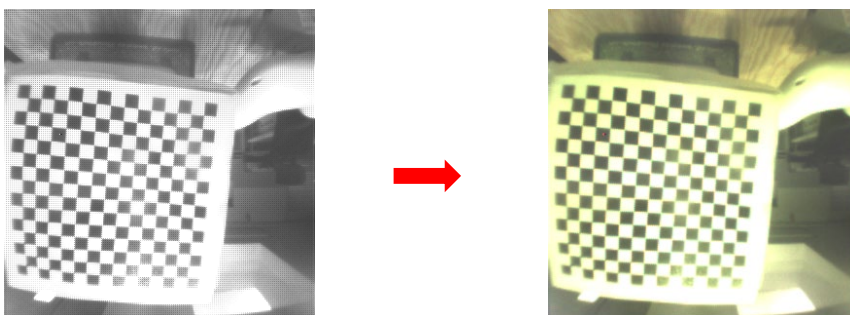
Vir: lasten.

2. Avtomatsko zajemanje slik s pomočjo programa

Kalibracijska tarča je osvetljena z reflektorjem, kar zagotavlja enako osvetlitev za vse kamere. Po premiku robotske roke se počaka na stabilizacijo, nato se zajame slika.

3. Obdelava vhodnih podatkov slik

Izvede se preslikava surove slike iz Bayerjevega formata BGGR v RGB z bilinearno interpolacijo, kar omogoča pravilno barvno predstavitev podatkov pred nadaljnjo analizo.

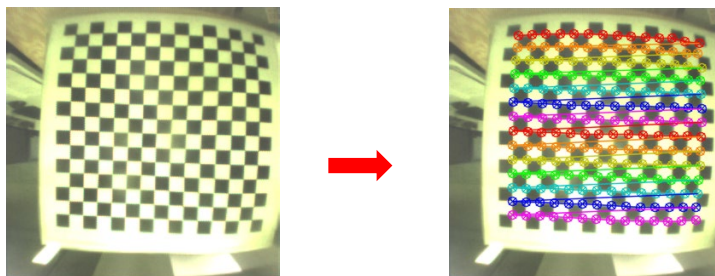


Slika 4: Preslikava surove slike v RGB

Vir: lasten.

4. Samodejno zaznavanje kalibracijske tarče s pomočjo računalniškega vida

Uporabljen je Förstner-jev operater (Forstner, 1987) za natančno detekcijo kotov šahovnice. Slike, pri katerih niso zaznani vsi koti kalibracijske tarče, se zavržejo.

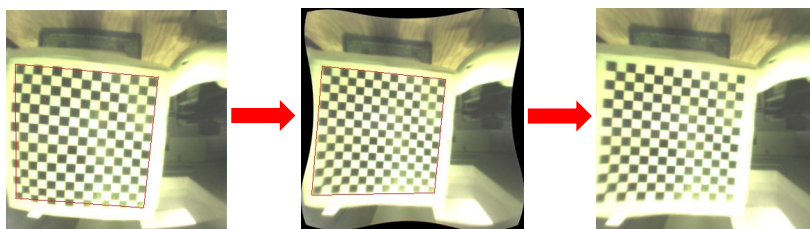


Slika 5: Detektirani koti šahovnice s Förstnerjevim operaterjem

Vir: lasten.

5. Optimizacija parametrov

Uporabljen je Levenberg-Marquardt (LM) (Gavin, 2019) algoritem za optimizacijo koeficientov tangencialne in radialne distorzije. Pri tem se minimizira napaka reprojekcije, kar pomeni, da se optimizirani parametri prilagajajo tako, da projekcija 3D točk na 2D slikovno ravnino najbolj ustreza dejansko zaznanim točkam na sliki.



Slika 6: Preslikava popačene slike v nepopačeno

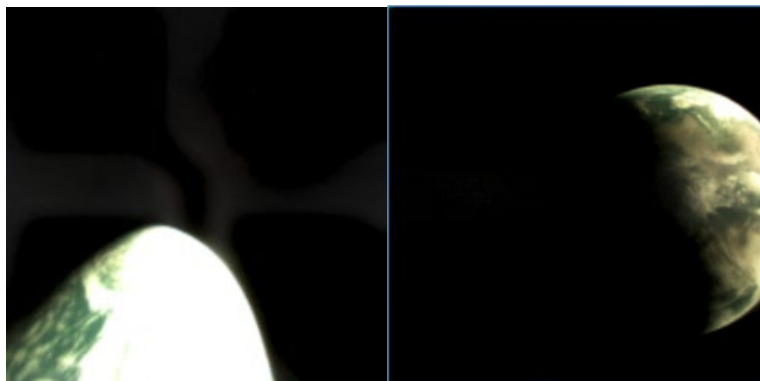
Vir: lasten.

$$\text{matrika kamere} = \begin{bmatrix} 248,15 & 0 & 156,97 \\ 0 & 245,30 & 153,91 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\text{koeficienti popačenja} = [0.45 \quad -2,45 \quad -0.01 \quad -0.01 \quad 2.22]$$

6. Validacija kalibracije

Validacija kalibracije se izvaja s kombinacijo avtomatizirane analize in človeškega pregleda. Python skripta izračuna reprojekcijsko napako, ki služi kot kvantitativni pokazatelj natančnosti kalibracije, medtem ko se vizualna potrditvena analiza uporablja za preverjanje pravilnosti poravnave in odprave optičnih distorzij.

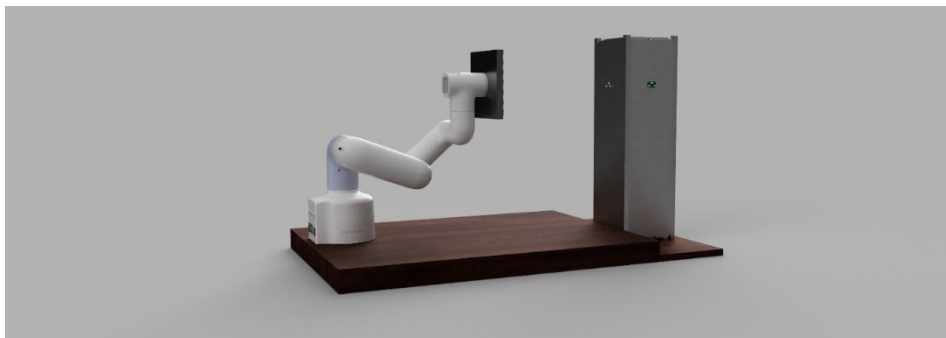


Slika 7: Zajeti sliki Zemlje z miniaturno kamero na satelitu TRISAT-R

Vir: lasten.

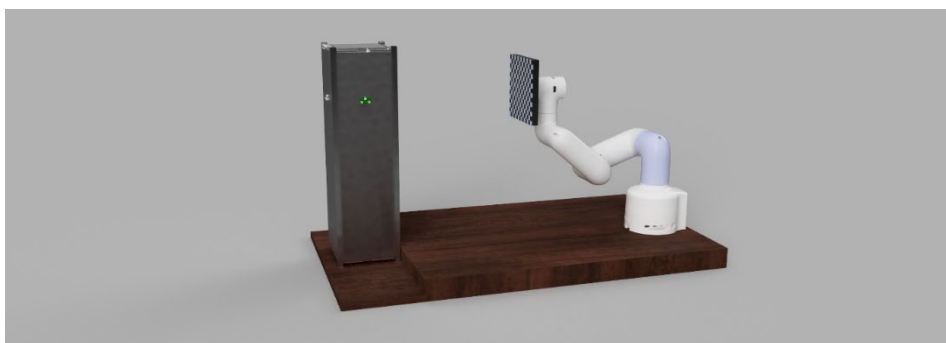
6 Koncept kalibracije satelitskih kamer in 360-stopinjske rekonstrukcije okolja

Primarni cilj satelitske misije TRISAT-S je tehnološka demonstracija varne radijske komunikacije med vesoljskim segmentom nanosatelita in zemeljsko postajo. Satelit pa bo imel na vsaki strani tudi integrirano miniaturno kamero. Kalibracija satelitskih kamer bo izvedena po zaključeni integraciji satelita, kar zahteva razvoj dodatnega sistema za rotacijo satelita okoli vseh treh osi. S tem bomo zagotovili enotne pogoje kalibracije za vse kamere ter omogočili popolnoma avtomatiziran proces kalibracije. V nadaljnjih raziskavah bomo analizirali možnost 360-stopinjske rekonstrukcije okolja z združevanjem slik, pridobljenih iz teh kamer. Cilj je razviti metodo za integracijo vidnih polj posameznih kamer v konsistentno panoramsko predstavitev prostora, pri čemer bo potrebno upoštevati geometrijsko poravnavo, popravke distorzije in združevanje slikovnih podatkov.



Slika 8: Računalniška upodobitev konceptne kalibracije kamer

Vir: lasten.



Slika 9: Računalniška upodobitev konceptne kalibracije kamer

Vir: lasten.

7 Zaključek

Kalibracija kamer je nepogrešljiv postopek v računalniškem vidu, saj omogoča korekcijo optičnih napak in zagotavlja natančnost rekonstrukcije 3D prostora. Avtomatizirani pristopi bistveno izboljšajo učinkovitost in ponovljivost kalibracije, zmanjšajo potrebo po človeškem posredovanju ter zagotavljajo visoko stopnjo natančnosti.

Prihodnje raziskave bodo osredotočene na rekonstrukcijo slik iz vseh zornih kotov v 360-stopinjski pogled. Sistem temelji na inženirskem modelu TRISAT-S satelita, ki bo opremljen s kamerami na vseh šestih straneh, kar omogoča zajem slik iz vseh smeri in njihovo natančno združevanje v celovito vizualno predstavitev.

Dodatne optimizacije kalibracijskega postopka s spremembami algoritmov ali pogojev niso načrtovane, saj trenutni sistem deluje odlično in dosega visoko stopnjo natančnosti. Nadaljnje raziskave bodo tako usmerjene predvsem v aplikacije obstoječega sistema in njegovo uporabo v kompleksnejših scenarijih računalniškega vida.

Viri in literatura

- Forstner, W. (1987). A fast operator for detection and precise location of distinct points, corners and center of circular features. In Proc. of the Intercommission Conference on Fast Processing of Photogrammetric Data, Interlaken, Switzerland, str 281–305, 1987.
- Gavin, H.P. (2019). The Levenberg-Marquardt algorithm for nonlinear least squares curve-fitting problems, Department of Civil and Environmental Engineering Duke University.
- Oosterwyck, V. N. (2018). Real Time Human Robot Interactions and Speed Control of a Robotic Arm for Collaborative Operations, DOI: 10.13140/RG.2.2.28723.53286.
- Sturm, P. (2014). Pinhole Camera Model. V: Ikeuchi, K. (eds) Computer Vision. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-31439-6_472.
- Zhang, Z. (2000). A Flexible New Technique for Camera Calibration, Pattern Analysis and Machine Intelligence, (str. 1330–1334).

EXPLORING THE DIFFERENCES IN PRUNING METHODS FOR CONVOLUTIONAL NEURAL NETWORKS

ROMANELA LAJIĆ, PETER PEER, ŽIGA EMERŠIČ

University of Ljubljana, Faculty of Computer and Information Science, Ljubljana,
Slovenia

romanela.lajic@fri.uni-lj.si, peter.peer@fri.uni-lj.si, ziga.emersic@fri.uni-lj.si

With the rising computational and memory cost of deep neural networks there is more effort to reduce the size of these models, especially when their deployment on resource constrained devices is the goal. New methods of compressing neural networks are being constantly developed with the goal of minimizing the drop in accuracy. In this paper we focus on pruning techniques as a way of compression. We present a comparison of different pruning criteria and analyze the loss in accuracy for the case of a simple non-iterative pruning procedure. We provide the comparison between cases when these criteria are applied to different architectures of convolutional neural networks.

DOI
[https://doi.org/
10.18690/um.feri.2.2025.2](https://doi.org/10.18690/um.feri.2.2025.2)

ISBN
978-961-286-960-1

Keywords:
convolutional neural
networks,
model compression,
model pruning,
deep learning,
deep neural networks



University of Maribor Press

DOI
[https://doi.org/
10.18690/um.feri.2.2025.2](https://doi.org/10.18690/um.feri.2.2025.2)

ISBN
978-961-286-960-1

Ključne besede:

konvolucijske nevronske mreže,
kompresija modelov,
obrezovanje modelov,
globoko učenje,
globoke nevronske mreže

RAZISKOVANJE RAZLIK MED METODAMI OBREZOVANJA ZA KONVOLUCIJSKE NEVRONSKE MREŽE

ROMANELA LAJČIĆ, PETER PEER, ŽIGA EMERŠIČ

Univerza v Ljubljani, Fakulteta za računalništvo in informatiko, Ljubljana, Slovenija
romanela.lajcic@fri.uni-lj.si, peter.peer@fri.uni-lj.si, ziga.emersic@fri.uni-lj.si

Zaradi naraščajočih računskih in pomnilniških zahtev globokih nevronskih mrež je vedno več truda usmerjenega v zmanjšanje velikosti teh modelov, še posebej kadar je cilj njihova uporaba na napravah z omejenimi viri. Nenehno se razvijajo nove metode za stiskanje nevronskih mrež, katerih cilj je čim manjši padec natančnosti. V tem članku se osredotočamo na tehnike obrezovanja (pruning) kot način stiskanja. Predstavimo primerjavo različnih kriterijev obrezovanja in analiziramo izgubo natančnosti pri enostavnem, neiterativnem postopku obrezovanja. Podamo primerjavo primerov, kjer so ti kriteriji uporabljeni pri različnih arhitekturah konvolucijskih nevronskih mrež.



Univerzitetna založba
Univerze v Mariboru

1 Introduction

Since the surge of popularity of deep neural networks in the area of computer vision has started, we have seen a growing trend when it comes to size and number of parameters of these models. This has also led to a rise in their memory and computational cost which makes their deployment on resource constrained devices, such as mobile or edge devices, challenging.

While there is a lot of effort invested in creating more lightweight models (Iandola, 2016; Buotros et al., 2022; Sandler et al., 2018), they can rarely achieve the accuracy comparable to the one of deep models. Compression attempts to reduce the number of parameters of larger models while maintaining the original accuracy or minimizing its reduction.

There are several different techniques of compressing neural networks. Some techniques such as quantization (Cai & Vasconcelos, 2020; Jacob et al., 2018; Zhou et al., 2017) focus on reducing the memory cost of a neural network by reducing the number of bits required for parameter representation. Other methods such as knowledge distillation (Hinton, 2015; Li et al., 2023; Park et al., 2019) attempt to train lightweight models so that they mimic the behavior of a larger architecture. Third group of methods, such as low-rank matrix decomposition (Lin et al., 2018) attempt to reduce computational cost of inference by reducing the number of operations in the network.

Pruning focuses on removing redundant connections or filters from a neural network. Connections which are considered redundant are the ones carrying either a low amount of information or less important information and are determined by a specific metric called the pruning criterion. The matter of choosing a pruning criterion is a topic of a large number of works. Criterion is mostly chosen in such a way that the reduction in the accuracy of the network is minimized, but can also be chosen in regards to a particular, specialized task such as reduction of bias in biometric models (Lin et al., 2022), improving discriminative power of the network (Liu et al., 2021) or enhancing generalization ability (Zimmer et al., 2023).

In this paper we will do a comparison and analysis of three different pruning criteria, based on different indicators in the neural network, such as outputs of layers, filter weights and batch normalization parameters. We will test the criteria on different architectures of convolutional networks and present results obtained on different datasets.

2 Related work

One of the early works presenting neural network pruning (LeCun et al., 1989) was published all the way back in 1989. The very idea for pruning stems from biology. At a young age children develop a large number of neural connections in order to make learning more efficient. Later on in life when a lot of the tasks learned in the earlier years become standard a lot of these connections are deactivated (Calderia et al., 2025). By the same logic, having a large number of layers and connections in a neural network makes information flow and learning easier, but during inference a lot of those connections are proven to be redundant and can be removed.

Pruning once again became popular with the development of deep learning, when the size of models began to significantly increase. At first, most of the work focused on finding the optimal pruning criterion, which would determine which of the connections can be removed from the network without a significant drop in accuracy. One of the first notable works around this time (Li et al., 2016) proposes computing L^1 norm of filter weights in each layer of a convolutional network. Filters with the lowest value of the norm are then removed under the assumption that they have the least effect on the output. Some methods focus on using the change in the loss function when a filter is pruned as an indicator of importance. Method presented in (You et al., 2019) proposes using a first-order Taylor expansion to estimate the change in the loss caused by setting a filter to zero. Certain papers such as (Liu et al., 2017) are based on the concept of sparsity regularization. The mentioned paper proposes using batch normalization layers as indicators and applies L^1 regularization to batch normalization weights before using their values to choose which filters to prune.

Several recent works exploring pruning focus on finding a new, more efficient pruning metric, which remains one of the biggest problems of the method. Shang et al. (Shang et al., 2022), propose breaking down the pruning procedure into layer-

level problems and solving them cooperatively. By assuming that the removal of a filter mostly affects the filters in the same layer, they propose using an evolutionary algorithm to choose a subpopulation of filters to keep for each layer. On the other hand, the work in (Liu et al., 2021) focuses more on improving the discriminative power of the network by introducing discriminative-aware losses such as cross-entropy to intermediate layers of the network, and combining them with feature-reconstruction error. Basha et al. (Basha et al., 2024) propose looking at the training history of the network. The hypothesis is that if the difference between the filters does not change significantly through the training epochs, those filters can be considered redundant. They suggest measuring the difference between the L^1 norms of different filter pairs during the training procedure and pruning one of the filters in the pairs with the smallest sums of absolute differences.

Li et al. (Li et al., 2022) argue that the structure of a network is just as important as the weights and that random channel pruning has the ability to reach performance levels of more complex pruning criterion. Although simply randomly choosing channels to prune cannot achieve competitive performance, the authors propose two setups based on random pruning. One is randomly choosing filters in a layer, then pruning them based on a certain criterion, such as L^1 . The other method is randomly choosing network configurations and training them in parallel.

Work in (Fang et al., 2023) addresses the issues with structural pruning. When performing structural pruning the architecture of the network is changed and interdependence between the parameters can oftentimes be violated. For this reason, the design often needs to be architecture specific. The authors attempt to find a way of automating structural pruning by representing the network as a graph and performing pruning by taking these dependencies into consideration.

Zimmer et al. (Zimmer et al., 2023) focuses on trying to enhance generalization ability of pruned networks by averaging out the parameters of different models. Since averaging the parameters of differently pruned models could actually increase sparsity of the final model, the authors propose so-called sparse model soups. This entails pretraining and pruning a larger model, then forming different models by changing other hyperparameters which can then be averaged. This allows the sparsity level to remain intact.

Some works have attempted to combine pruning with other compression techniques. In (Li et al., 2023) pruning is combined with mixed-precision quantization for a more efficient hardware acceleration. Iterative quantization is performed until the redundant weights are completely pruned and the rest of the network is quantized with a different bit-width.

3 Methodology

Pruning of filters in convolutional neural network in most cases consists of three steps: choosing the least important filters which can be removed, removing the filters and fine-tuning the model. The metric which implies which filters are the least important is called the pruning criterion, while the percentage of the filters chosen to be removed is referred to as pruning sparsity. We will be examining three different pruning criteria and comparing their performance.

First examined algorithm is based on the output of the filters. We consider a greedy algorithm which chooses filters based on their output norm. The algorithm removes filters one by one by choosing the filter with the lowest output norm after the removal of the previous filter until we achieve the desired sparsity.

Other two methods determine redundant filters based on network parameters. The first method determines the least important filters by calculating their L^1 norm under the assumption that the filters with the lowest norm contribute less to the output. The second method, proposed in (Liu et al., 2017), looks at weights of a batch normalization layer which is typically placed after the convolutional layer. The method applies L^1 regularization to batch normalization layers, after which it chooses the filters to prune based on the corresponding batch norm weights.

We evaluate the three techniques applied to three architectures of convolutional networks, VGG-16, ResNet-18 and ResNet-50, on CIFAR-10 and CIFAR-100 datasets. We also provide results of the VGG-16 network on the ImageNet dataset.

4 Experiments and Results

When it comes to CIFAR datasets, all three networks have been trained from scratch after which pruning has been applied using the three described criteria with the same sparsity level of 40%. Pruned networks are then retrained on epochs, with a batch

size 32, using an SGD optimizer starting at a learning rate of $1e-3$ which is then reduced to values $1e-4$ and after that $1e-5$. When using the CIFAR-10 dataset, the models are retrained on 10 epochs, and on the CIFAR-100 dataset, on 20 epochs. Random rotation and random horizontal flip are applied to the training images. The results are shown in the Table 1 and Table 2. When it comes to the slimming technique the original paper applies L^1 regularization to the batch normalization layers during training. Considering that the tests on the ImageNet dataset are done using a pretrained model, the regularization is omitted in the other two datasets as well. This kind of test will give us an idea of how well batch normalization weights function as an indicator of importance on their own, without any additional preparation, which might prove to be useful in cases where training the original model from scratch is simply not possible for various reasons.

Table 1: Results on CIFAR-10

Model	Method	Accuracy
VGG-16	Pre-pruning	93.18%
	Greedy	91.01%
	L1	90.67%
	Slimming	90.66%
ResNet-18	Pre-pruning	94.41%
	Greedy	93.37%
	L1	93.38%
	Slimming	93.54%
ResNet-50	Pre-pruning	94.96%
	Greedy	93.69%
	L1	93.99%
	Slimming	92.63%

From the tables we can see that on most combinations of architecture and dataset the pruning criteria give comparable results. There is also a notable drop in performance in most of the cases, in some being more prominent than the others. The results imply that the drop in performance of the pruned models is affected by many factors, including the original architecture, the dataset and the retraining procedure.

VGG-16 is also tested on the ImageNet dataset. Retraining a pruned network on ImageNet is a more challenging task and some works use iterative retraining procedures where the network is retrained after pruning each of the layers, in order to get minimize the loss of accuracy as much as possible. For the sake of direct

comparisson, the network is retrained on ImageNet in the similar way as on CIFAR datasets. The network is retrained on 10 epochs after completely pruning all of the convolutional layers, using the same parameters as described for the CIFAR datasets. The training images are center-cropped to the appropriate size and random horizontal flip is applied. No other augmentations are added to the images. The results are shown in the Table 3.

Table 2: Results on CIFAR-100

Model	Method	Accuracy
VGG-16	Pre-pruning	71.50%
	Greedy	68.26%
	L1	67.26%
	Slimming	67.92%
ResNet-18	Pre-pruning	75.72%
	Greedy	74.11%
	L1	75.08%
	Slimming	74.95%
ResNet-50	Pre-pruning	78.76%
	Greedy	76.87%
	L1	75.66%
	Slimming	77.40%

Table 3: Results on ImageNet

Model	Method	Accuracy
VGG-16	Pre-pruning	73.42%
	Greedy	68.39%
	L1	68.77%
	Slimming	68.50%

In the case of ImageNet there is also a comparable performance between the different pruning criteria. Performance loss in this case is more significant, since complex datasets require more complex procedures, such as iterative retraining after pruning each of the layers, as was previously mentioned. This allows the network to gradually adjust to the loss of information.

5 Conclusion

In this paper we have explored different pruning criteria based on both outputs of layers and network parameters. We have tested and compared these criteria on three different architectures of convolutional neural networks using three different

datasets. From the results we can conclude that the metrics give comparable results for most cases but that there is a certain drop in accuracy when networks are reduced to 60% of their convolutional filters. In order for the pruned models to give accuracy which matches the one of the full-sized model, more complex retraining procedures must be applied even for simple datasets, while for more challenging cases both more complex training procedures and more advanced pruning criteria.

References

- Basha, S. S., Farazuddin, M., Pulabaigari, V., Dubey, S. R., & Mukherjee, S. (2024). Deep model compression based on the training history. *Neurocomputing*, 573, 127257.
- Boutros, F., Siebke, P., Klemm, M., Damer, N., Kirchbuchner, F., & Kuijper, A. (2022). Pocketnet: Extreme lightweight face recognition network using neural architecture search and multistep knowledge distillation. *IEEE Access*, 10, 46823-46833.
- Cai, Z., & Vasconcelos, N. (2020). Rethinking differentiable search for mixed-precision neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 2349-2358).
- Caldeira, E., Neto, P. C., Huber, M., Damer, N., & Sequeira, A. F. (2025). Model compression techniques in biometrics applications: A survey. *Information Fusion*, 114, 102657.
- Fang, G., Ma, X., Song, M., Mi, M. B., & Wang, X. (2023). Depgraph: Towards any structural pruning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 16091-16101).
- Hinton, G. (2015). Distilling the Knowledge in a Neural Network. *arXiv preprint arXiv:1503.02531*.
- Iandola, F. N. (2016). SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size. *arXiv preprint arXiv:1602.07360*.
- Jacob, B., Kligys, S., Chen, B., Zhu, M., Tang, M., Howard, A., Adam, H., Kalenichenko, D. (2018). Quantization and training of neural networks for efficient integer-arithmetic-only inference. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2704-2713).
- LeCun, Y., Denker, J., & Solla, S. (1989). Optimal brain damage. *Advances in neural information processing systems*, 2.
- Li, H., Kadav, A., Durdanovic, I., Samet, H., & Graf, H. P. (2016). Pruning filters for efficient convnets. *arXiv preprint arXiv:1608.08710*.
- Li, J., Guo, Z., Li, H., Han, S., Baek, J. W., Yang, M., ... & Suh, S. (2023). Rethinking feature-based knowledge distillation for face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 20156-20165).
- Li, Y., Adamczewski, K., Li, W., Gu, S., Timofte, R., & Van Gool, L. (2022). Revisiting random channel pruning for neural network compression. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 191-201).
- Liu, J., Zhuang, B., Zhuang, Z., Guo, Y., Huang, J., Zhu, J., & Tan, M. (2021). Discrimination-aware network pruning for deep model compression. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(8), 4035-4051.
- Lin, S., Ji, R., Chen, C., Tao, D., & Luo, J. (2018). Holistic cnn compression via low-rank decomposition with knowledge transfer. *IEEE transactions on pattern analysis and machine intelligence*, 41(12), 2889-2905.
- Lin, X., Kim, S., & Joo, J. (2022, October). Fairgrape: Fairness-aware gradient pruning method for face attribute classification. In *European Conference on Computer Vision* (pp. 414-432). Cham: Springer Nature Switzerland.

- Liu, Z., Li, J., Shen, Z., Huang, G., Yan, S., & Zhang, C. (2017). Learning efficient convolutional networks through network slimming. In *Proceedings of the IEEE international conference on computer vision* (pp. 2736-2744).
- Li, Z., Gong, Y., Zhang, Z., Xue, X., Chen, T., Liang, Y., ... & Wang, Z. (2023). Accelerable lottery tickets with the mixed-precision quantization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 4604-4612).
- Park, W., Kim, D., Lu, Y., & Cho, M. (2019). Relational knowledge distillation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 3967-3976).
- Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4510-4520).
- Shang, H., Wu, J. L., Hong, W., & Qian, C. (2022). Neural network pruning by cooperative coevolution. *arXiv preprint arXiv:2204.05639*.
- You, Z., Yan, K., Ye, J., Ma, M., & Wang, P. (2019). Gate decorator: Global filter pruning method for accelerating deep convolutional neural networks. *Advances in neural information processing systems*, 32.
- Zhou, S. C., Wang, Y. Z., Wen, H., He, Q. Y., & Zou, Y. H. (2017). Balanced quantization: An effective and efficient approach to quantized neural networks. *Journal of Computer Science and Technology*, 32, 667-682.
- Zimmer, M., Spiegel, C., & Pokutta, S. (2023). Sparse model soups: A recipe for improved pruning via model averaging. *arXiv preprint arXiv:2306.16788*.

GENERIRANJE SINTETIČNIH CT SLIK IZ MR SLIK PODROČJA GLAVE IN VRATU Z UPORABO DIFUZIJSKIH MODELOV

ROK MARKO ŠTER, GAŠPER PODOBNIK, TOMAŽ VRTOVEC

Univerza v Ljubljani, Fakulteta za elektrotehniko, Ljubljana, Slovenija
rs7839@student.uni-lj.si, gasper.podobnik@fe.uni-lj.si, tomaz.vrtovec@fe.uni-lj.si

Računalniška tomografija (CT) je slikovna preiskava, ki se v klinični praksi standardno zajame v okviru načrtovanje radioterapije. V primeru raka območju glave in vratu (HaN) se pogosto zajame tudi magnetno resonančne (MR) slike za natančnejše orisovanje tumorjev in kritičnih organov. V zadnjem času se vse bolj uveljavlja radioterapija na podlagi MR-samostojnega pristopa, ki odstrani potrebo po zajemu CT slik in s tem izpostavljenost ionizirajočemu sevanju, vendar pa zahteva rešitev za generiranje sintetičnih CT slik na podlagi MR slik. Nedavne študije kažejo, da difuzijski modeli nudijo realistično generiranje slik z natančnimi anatomskimi podrobnostmi in manj artefakti kot generativne nasprotniške mreže. V tej študiji smo razvili difuzijski model za pretvorbo MR slik v sintetične CT slike za HaN področje. Naš pristop, ovrednoten na zbirki podatkov HaN-Seg, ki vključuje pare CT in MR slik istih bolnikov, doseže indeks strukturne podobnosti 92,2 %, vršno razmerje signal-šum 33,1 dB ter povprečno absolutno napako 35,3 HU. Model dodatno ovrednotimo tudi za segmentacijo kritičnih organov. Rezultati potrjujejo potencial uporabe difuzijskih modelov pri načrtovanju radioterapije.

DOI
[https://doi.org/
10.18690/um.feri.2.2025.3](https://doi.org/10.18690/um.feri.2.2025.3)

ISBN
978-961-286-960-1

Ključne besede:
radioterapija,
računalniška tomografija,
magnetna resonanca,
difuzijski modeli,
difuzija,
umetna inteligenca,
globoko učenje,
sintetične slike,
računalniško podprta
analiza medicinskih slik

DOI
[https://doi.org/
10.18690/um.feri.2.2025.3](https://doi.org/10.18690/um.feri.2.2025.3)

ISBN
978-961-286-960-1

Keywords:

radiotherapy,
computed tomography,
magnetic resonance,
diffusion models,
diffusion,
artificial intelligence,
deep learning,
synthetic images,
computer-assisted medical
image analysis

GENERATION OF SYNTHETIC CT IMAGES FROM MR IMAGES IN THE HEAD AND NECK REGION USING DIFFUSION MODELS

ROK MARKO ŠTER, GAŠPER PODOBNIK, TOMAŽ VRTOVEC

University of Ljubljana, Faculty of Electrical Engineering, Ljubljana, Slovenija
rs7839@student.uni-lj.si, gasper.podobnik@fe.uni-lj.si, tomaz.vrtovec@fe.uni-lj.si

In cancer radiotherapy, CT images are essential for planning, while MR images accurately delineate tumors and organs-at-risk, especially in the head and neck (HaN) region. MR-only radiotherapy, with which we generate synthetic CT images from MR data, removes patient radiation exposure. Recent studies indicate that diffusion models yield more realistic images with precise anatomical details and fewer artifacts than generative adversarial networks. In this study, we employ a diffusion model to translate MR images into synthetic CT images for the HaN region. Evaluated on the HaN-Seg dataset of paired CT and MR images of the same patients, our approach achieves a structural similarity index of 92.2%, a peak signal-to-noise ratio of 33.1 dB, and a mean absolute error of 35.3 HU, demonstrating its potential in radiotherapy planning. Validation was extended on a downstream task of organ-at-risk segmentation. Results demonstrate the potential of applying diffusion models into the radiotherapy workflow.



1 Uvod

Radioterapija raka na področju glave in vratu (HaN) zahteva natančno lokalizacijo tumorskega tkiva in kritičnih organov (OAR) na podlagi slik za načrtovanje obsevalnega zdravljenja, kar omogoča načrtovanje obsevanja z minimalnim vplivom na zdrava tkiva (Gregoire in dr., 2020). Računalniško tomografske (CT) slike so tradicionalno temelj za takšno načrtovanje, saj vsebujejo informacije o atenuaciji v tkivih, ki so ključne za izračun doznega načrta. V primerjavi s CT ponuja magnetna resonanca (MR) boljši kontrast mehkega tkiva, zaradi česar je neprecenljiva za natančno orisovanje meja tumorjev in OAR, še posebej v zapletenem območju HaN (Podobnik in dr., 2023). Ker je slikanje s CT povezano z ionizirajočim sevanjem (Podobnik in dr., 2024a), se je razvila ideja o tako imenovanem MR-samostojnem (MR-only) načrtovanju obsevanja, ki v celoti sloni le na uporabi MR slik (McDonald in dr., 2024). Ker izračun doznega načrta temelji na informaciji o elektronski gostoti, ki ni enostavno izračunljiva iz MR slike, se je uveljavila ideja o generiranju sintetičnih CT slik na podlagi MR slik, za kar se je uveljavil tudi izraz prevajanje med slikovnimi modalitetama (angl. image-to-image translation, I2I).

Primarni cilj te študije je implementirati in oceniti nov pristop za omenjeni problem, ki temelji na difuzijskih modelih, ki so se v zadnjem času uveljavili kot uporabna metoda globokega učenja. Na podlagi pregleda literature ugotavljamo, da difuzijski modeli predstavljajo znaten napredek v primerjavi s tradicionalnimi generativnimi nasprotniškimi mrežami, saj ponujajo bolj stabilen postopek učenja ter izboljšano realističnost slik in anatomsko natančnost z manj artefakti (Dayarathna in dr., 2024). V tej študiji se osredotočamo na primerjavo dveh različnih strategij I2I: 2D pristop generiranja 2D aksialnih rezin in 3D pristop generiranja manjših 3D koščkov slike. Uporabljamo javno objavljeno zbirko podatkov, ki izhaja iz Onkološkega inštituta Ljubljana: "HaN-Seg: The Head and Neck Organ-at-Risk CT and MR Segmentation Dataset" (Podobnik in dr., 2023). Nabor podatkov vključuje pare CT in MR slik istih bolnikov skupaj s referenčnimi segmentacijami OAR, ki so jih strokovni orisovalci ustvarili ročno. MR slike uporabljamo kot pogoj za generiranje sintetičnih CT (sCT) slik, ki se nato ocenjujejo glede na anatomsko natančnost v primerjavi z dejanskimi ciljnim CT slikami ter celotno kakovostjo slike. Realističnost sCT slik ocenimo na nalogi segmentacije slik, in sicer tako, da primerjamo referenčne segmentacije s segmentacijami, ki jih ustvari orodje TotalSegmentator (Wasserthal in dr., 2023), naučeno izključno na realnih CT slikah iz podatkovnih množic, ki ne vključujejo

zbirke HaN-Seg. V okviru analize primerjamo uspešnost segmentacije na realnih CT slikah iz baze HaN-Seg z rezultati segmentacije na generiranih sCT slikah. Obravnavani pristop je zanimiv z večih vidikov: I2I na podlagi MR slik je ključen gradnik MR-samostojnega načrtovanja radioterapije, ki izniči bolnikovo izpostavljenost sevanju tekom načrtovanja radioterapije in ponuja stroškovno učinkovito alternativo tradicionalnim pristopom slikanja z več modalitetami.

2 Metode

2.1 Zbirka slik

V tej študiji smo uporabili nabor podatkov HaN-Seg (Podobnik in dr., 2023), ki ga sestavljajo 3D pari CT in MR slik 56 oseb. Vsaka oseba je prestala CT in T1w MR slikanje na Onkološkem inštitutu v Ljubljani z namenom zdravljenja raka na področju HaN. Uporabili smo uradno razdelitev slikanih oseb na učno in testno množico, z 42 pari slik v učni in 14 pari slik v testni množici, ter poročali o rezultatih, izmerjenih samo na testni množici, s čimer poročani rezultati služijo kot merilo za primerjavo orodij za I2I na tej javno dostopni zbirki slik.

2.2 Predobdelava slik

Vsak par CT in MR slik je bil poravnan s kombinacijo uporabe toge in elastične poravnave, kot je podrobno opisano v študiji o variabilnosti opazovalcev, izvedeni na naboru podatkov HaN-Seg (Podobnik in dr., 2024a). Intenzitete CT slik so bile linearno oknjene med intenzitetama -1000 in 3000 HU ter nato preslikane na interval $[-1, 1]$. Čeprav preslikava vrednosti izven intervala na mejne vrednosti vodi do izgube dela informacij, je ta izguba zanemarljiva, saj odrezane intenzitete primarno prikazujejo artefakte ali šum v ozadju, ki ni primarnega pomena za našo analizo. Intenzitete MR slik smo linearno preslikali s tako imenovano *z-normalizacijo* (tj. na ničelno povprečje in enotino varianco), saj smo opazili, da takšna normalizacija bolje deluje za MR slike zaradi prisotnosti artefaktov zobnih implantatov, ki povzročajo intenzitete ekstremnih vrednosti in posledično izkrivljajo porazdelitev intenzitet v primeru uporabe normalizacije intervalov. Poravnane in normalizirane CT in MR slike so bile nato ponovno vzorčene na velikost vokslov $1 \times 1 \times 1$ mm³ in središčno obrezane v aksialni ravnini na mrežo velikosti $256 \times 256 \times Z$ slikovnih

elementov, kjer Z predstavlja število aksialnih rezin. Pomembno je omeniti, da se Z med slikami razlikuje zaradi razlik v vidnem polju originalnih slik.

2.3 Verjetnostni modeli difuzijskega odstranjevanja šuma

Uporabili smo difuzijski model, ki temelji na arhitekturi U-Net (Ronneberger in dr., 2015), razširjen z mehanizmi vgrajevanja časovnih korakov in pozornosti, ter razporejevalnik dodajanja šuma DDPM (angl. denoising diffusion probabilistic model) (Ho in dr., 2020), ki sta implementirana v knjižnici MONAI¹ Generative Models² (Cardoso in dr., 2022). Med učenjem model sprejme vhodno sliko šuma in sliko, s katero pogojujemo model pogoja. V našem primeru je to MR slika, ki vodi proces difuzije tako, da je sCT slika v anatomskem ujemanju z MR sliko. V vsakem koraku učenja modela se vsaki sliki CT v paketu učenja dodeli naključni celoštevilski časovni korak. Časovni koraki se vzorčijo iz enakomerne porazdelitve med 0 in N_{train, T_s} ter se uporabljajo za simulacijo različnih količin šuma, dodanega sliki. Vsak učni vzorec v paketu je sestavljen iz originalne CT slike z dodanim Gaussovimi šumom na podlagi naključnega časovnega koraka in načina delovanja razporejevalnika dodajanja šuma. Model se nato uči napovedovanja dodanega šuma na podlagi slike s šumom in MR slike. Čeprav obstaja veliko strategij za posredovanje informacij modelu, smo izvedli združitev slike šuma in MR slike na nivoju slikovnega kanala. Informacije pogojne slike, tj. MR, služijo kot vodilo za model, saj zagotavljajo informacije o anatomski strukturi in podrobnostih mehkega tkiva (Kazerouni in dr., 2023). Med inferenco se uporablja iterativni postopek odstranjevanja šuma: vhod modela sta naključni šum in MR slika. Ta vhod se nato v iterativnem postopku odstranjevanja šuma obdeluje z odštevanjem napovedane slike šuma, kar na koncu privede do realistične sCT slike, ki je anatomsko podobna vhodni MR sliki.

3 Eksperimenti in rezultati

3.1 Eksperimenti generiranja sintetičnih CT slik

Zasnovali smo dva poskusa, pri katerih smo zagotovili čim bolj dosledno strategijo učenja in nastavitve modela, hkrati pa smo omogočili primerjavo med pristopoma 2D generiranja rezin in 3D generiranja koščkov. Za difuzijski model smo uporabili

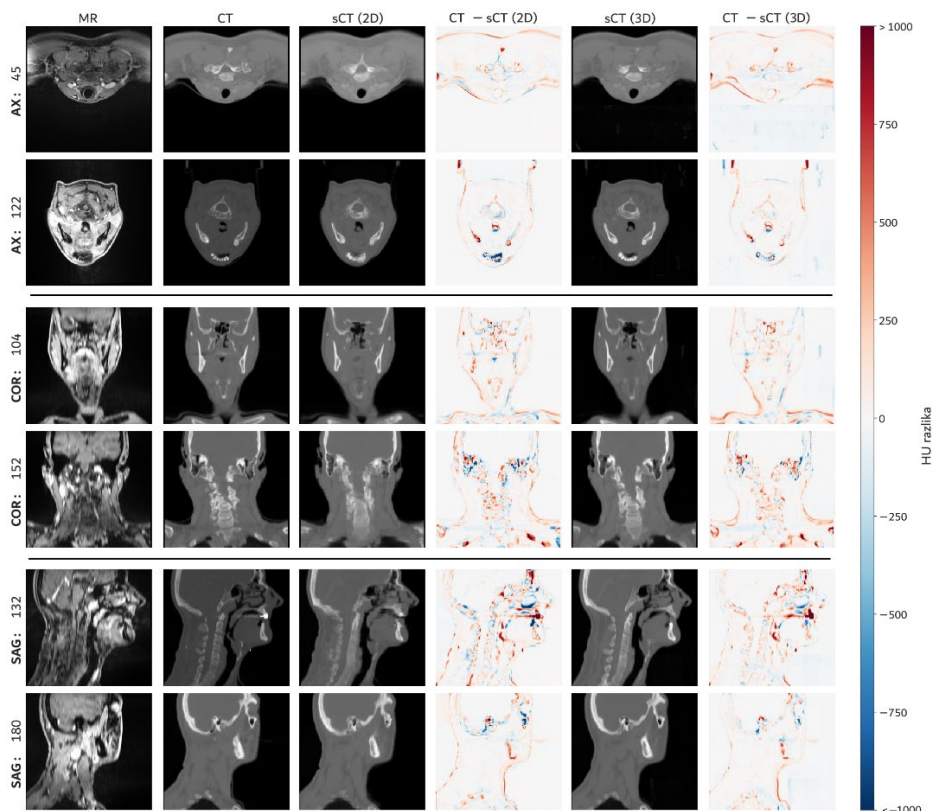
¹ <https://monai.io/>

² <https://github.com/Project-MONAI/GenerativeModels>

arhitekturo U-Net, ki vključuje vgradnjo časovnega koraka in mehanizem tako imenovane *samopozornosti* v dveh najglobljih plasteh mreže. Model 2D je bil naučen z aksialnimi rezinami velikosti 256×256 slikovnih elementov z uporabo mreže U-Net z [32, 64, 128, 256, 512] kanali značilnic, medtem ko je bil 3D model naučen s koščki velikosti $64 \times 64 \times 64$ slikovnih elementov z uporabo mreže U-Net s [64, 128, 256, 512] kanali značilnic. Pri uporabi 2D modela so bile generirane aksialne rezine kasneje enostavno zložene v pravilnem vrstnem redu za namen rekonstrukcije celotne 3D sCT slike. Uporabili smo kombinacijo kriterijskih funkcij L1 in srednje kvadratne napake (MSE) z empirično določenima utežema 0,75 in 0,25. Model smo učili z optimizatorjem AdamW z začetno vrednostjo učene konstante $1e-4$ in z razporejevalnikom dodajanja šuma DDPM z $N_{\text{train}_T_s} = 1000$ časovnimi koraki ter njegovimi sicer privzetimi parametri MONAI modula. V primerjavi z difuzijskimi implicitnimi modeli odstranjevanja šuma (Song in dr., 2021) in psevdonumeričnimi difuzijskimi modeli (Liu in dr., 2022) je razporejevalnik DDPM generiral najbolj anatomsko pravilne in smiselne rezultate za enako število časovnih korakov in epoh. Empirično smo ugotovili, da tehnike bogatenja podatkov, kot so rotacija, skaliranje, translacija in gama transformacija, niso izboljšale učinkovitosti modela, zato med učenjem končnih modelov tehnike bogatenja podatkov nismo uporabili. Med učenjem modela smo redno merili kvantitativne metrike na validacijski³. Ker je celotna inferenca z difuzijskimi modeli razmeroma računsko potratna, smo izvedli dve vrsti vrednotenja med učenjem, ki ju imenujemo enostavno in celovito vrednotenje. Enostavno vrednotenje je bilo izvedeno vsakih 25 epoh, ko smo preprosto izmerili vrednost kriterijske funkcije na slikah testne množice v naključno izbranih časovnih korakih, medtem ko je bilo celovito vrednotenje izvedeno vsakih 100 epoh in je vključevalo celoten postopek obratnega difuzijskega procesa (odstranjevanje šuma), in sicer z reduciranim številom časovnih korakov na 100. Končni model za vsak pristop je bil izbran na podlagi najnižje vrednosti kriterijske funkcije, izračunane pri enostavnem vrednotenju. To je pomenilo skupaj 7.800 in 29.700 epoh, ki so trajale približno 2 oziroma 4 dni za pristop z 2D rezinami in s 3D koščki. Modele z nalogo napovedi 2D aksialnih rezin smo učili na delovni postaji, opremljeni z grafično procesno enoto (GPU) Nvidia RTX 2080Ti z 12 GB pomnilnika, 3D modele pa je bilo treba učiti na GPU Nvidia RTX A6000 z 48 GB pomnilnika.

³ Dve sliki iz originalne učnega množice sta bili premaknjeni v tako imenovano validacijsko množico.

3.2 Vrednotenje sintetičnih CT slik



Slika 1: Primeri sintetičnih CT slik, ustvarjenih z našimi 2D in 3D modeli, pri katerih je število sklepanja nastavljeno na 100 na podlagi ustreznih vhodnih MR slik. V prvih dveh vrsticah so prikazani aksialni prerezi, v naslednjih dveh vrsticah koronalni prerezi, v spodnjih dveh vrsticah pa sagitalni prerezi. Poleg vsake ustvarjene slike je na voljo grafični prikaz razlik med pravo in sintetično CT sliko, pri čemer so razlike v Hounsfieldovih enotah označene z rdečo in modro barvo v skladu z legendo na skrajni desni strani.

Vir: lasten.

Dobljeni modeli so bili vrednoteni na 14 parih 3D slik iz testne množice z uporabo treh kvantitativnih metrik: vršno razmerje signal-šum (PSNR), indeks strukturne podobnosti (SSIM) (Wang in dr., 2004) in srednja absolutna napaka (MAE). PSNR in SSIM smo izračunali tako v 2D, tj. primerjava sintetičnih aksialnih rezin s pripadajočimi ciljnimi aksialnimi rezinami, kot v tudi 3D. V nadaljevanju se okrajšavi PSNR in SSIM nanašata na njun 3D izračun, ki je običajen izračun na področju

obdelave 3D slik. Rezultati so navedeni v tabeli 1, na sliki 1 pa so prikazani primeri sCT.

Tabela 1: Metrike PSNR, SSIM in MAE pri generiranju sCT slik z 2D in 3D modeli.

Model	PSNR-2D (dB)	PSNR-3D (dB)	SSIM-2D (%)	SSIM-3D (%)	MAE (HU)
2D	36,9 ± 2,7	32,1 ± 1,2	90,8 ± 2,1	90,3 ± 2,1	33,1 ± 4,9
3D	35,1 ± 1,4	33,1 ± 1,4	93,0 ± 1,4	92,2 ± 1,4	35,3 ± 4,7

3.3 Vrednotenje uporabnosti sintetičnih CT slik

Pomembno merilo uporabnosti sintetičnih slik je njihova realističnost. Medtem ko kvantitativne metrike podobnosti služijo kot učinkovito orodje za primerjavo raznolikih metod, se ob tem pojavijo njihove specifične omejitve. Čeprav vizualno vrednotenje kakovosti ponuja celovitejši vpogled, temelji na subjektivnih ocenah opazovalcev ter je časovno in, v primeru uporabe strokovnjakov iz klinične prakse, tudi precej finančno potratno. V okviru prispevka smo zato uporabili alternativen pristop, kjer smo uspešnost segmentacije OAR na sCT slikah ocenili s pomočjo avtomatske metode, naučene izključno na realnih CT slikah, pridobljenih iz drugih podatkovnih zbirk. Za ta namen smo uporabili segmentacijsko orodje TotalSegmentator in z njim pridobili segmentacije osmih anatomskih struktur: levo oko, desno oko, levi optični živec, desni optični živec, leva parotidna žleza, desna parotidna žleza, hrbtenjača ter ščitnica. Rezultate segmentacije na sCT slikah smo nato primerjali z rezultati, doseženimi na realnih CT slikah.

Uspešnost segmentacije orodja TotalSegmentator smo ocenili z dvema metrikama: Diceov koeficient podobnosti (DSC) in 95. percentil Hausdorffove razdalje (HD95). Kot referenco smo uporabili ročne segmentacije iz uporabljene podatkovne zbirke HaN-Seg, kjer smo za vsako zgoraj navedeno strukturo izračunali razliko metrik, izračunanih na segmentacijah realnih in sintetičnih slik na naslednji način:

$$\Delta DSC = DSC_{sCT} - DSC_{CT}, \quad (1)$$

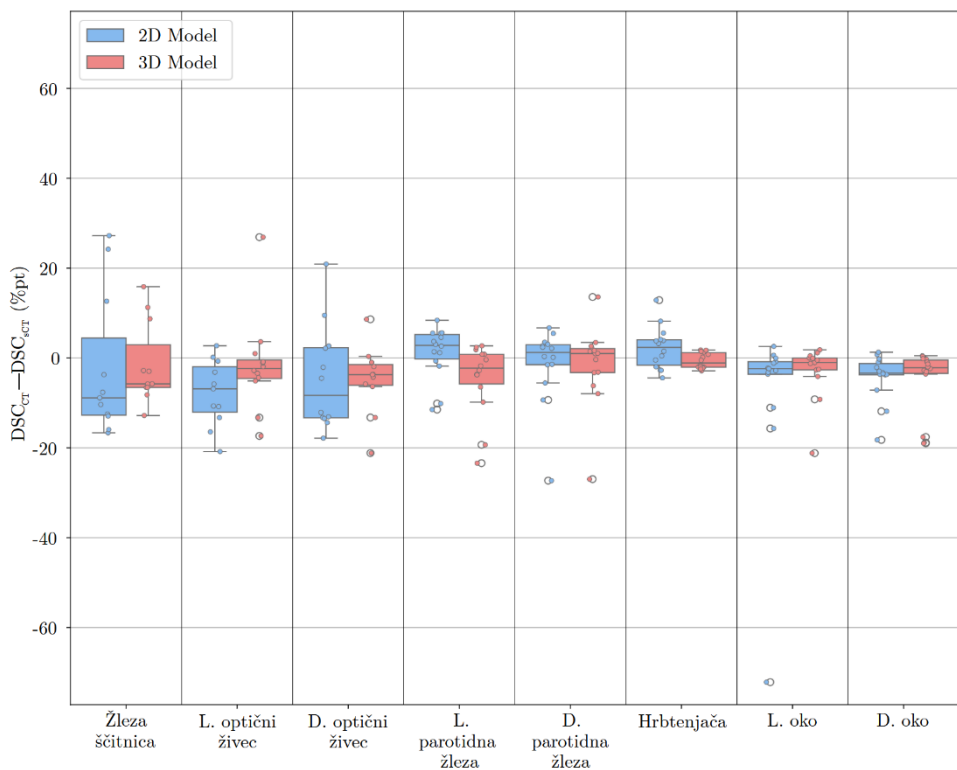
kjer je DSC_{sCT} vrednost DSC med referenčno segmentacijsko masko anatomske strukture in segmentacijsko masko iste strukture, ustvarjeno z orodjem TotalSegmentator na podlagi sCT slike. Vrednost DSC_{CT} pa predstavlja DSC med

referenčno segmentacijsko masko anatomske strukture in segmentacijsko masko, ustvarjeno z orodjem TotalSegmentator na podlagi realne CT slike.

Izračun $\Delta HD95$ poteka na podoben način:

$$\Delta HD95 = HD95_{sCT} - HD95_{CT}, \quad (2)$$

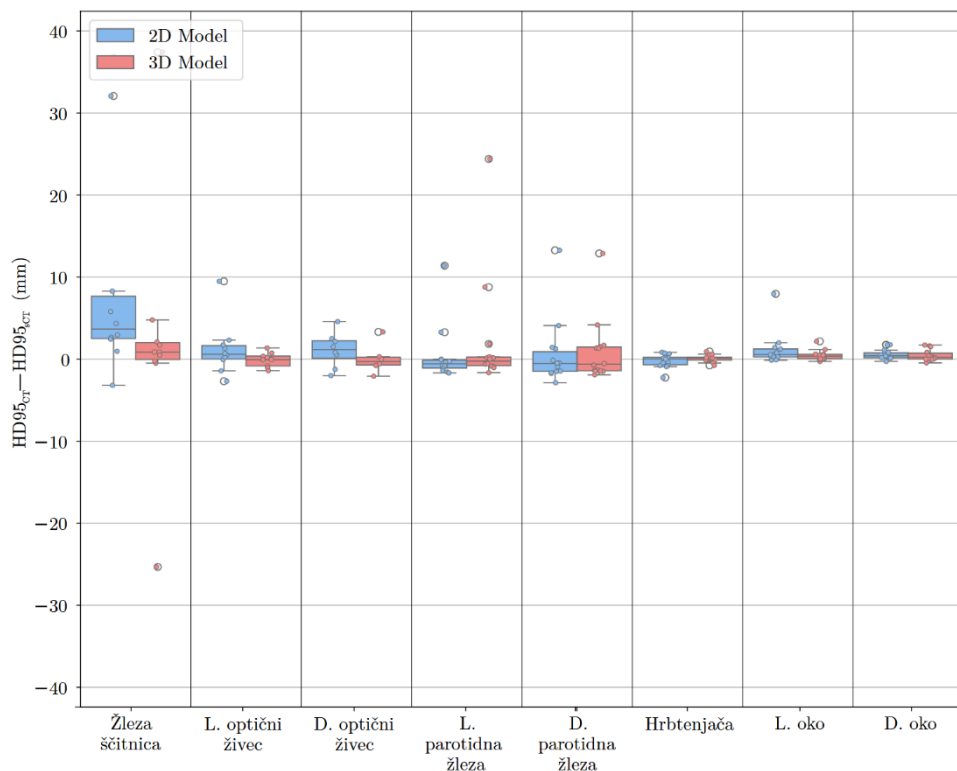
kjer gre za enak princip računanja $\Delta HD95$ kot pri izračunu ΔDSC , oznaki $HD95_{sCT}$ in $HD95_{CT}$ pa sledita istemu principu kot DSC_{sCT} in DSC_{CT} .



Slika 2: Grafični prikaz razlike v vrednosti DSC na segmentacijah sintetičnih in realnih slik, primerjanih z referenčnimi orisi. Barva na škatlastem diagramu prikazuje, kateri model smo uporabili za generiranje sintetičnih slik: modra predstavlja 2D model, rdeča pa 3D model.

Vir: lasten.

Rezultati te analize so prikazani na škatlastih diagramih na slikah 2 in 3. Pozitivne vrednosti Δ DSC pomenijo, da so segmentacije na podlagi sCT slik boljše, tj. bolj podobne ročnim segmentacijam, enako velja za negativne vrednosti Δ HD95. Modra barva označuje metrike 2D modela, medtem ko rdeča barva označuje metrike 3D modela. Vse primere, kjer orodje TotalSegmentator ni uspelo segmentirati anatomskih struktur, smo odstranili iz nadaljnje analize.



Slika 3: Grafični prikaz razlike v vrednosti HD95 na segmentacijah sintetičnih in realnih slik, primerjanih z referenčnimi orisi. Barva na škatlastem diagramu prikazuje, kateri model smo uporabili za generiranje sintetičnih slik: modra predstavlja 2D model, rdeča pa 3D model.

Vir: lasten.

4 Diskusija

Ta študija je prva, ki poroča o rezultatih I2I na javno dostopnem naboru podatkov HaN-Seg (Podobnik in dr., 2023). Čeprav je bil ta nabor podatkov objavljen predvsem za primerjavo algoritmov za segmentacijo OAR iz slik več modalitet na

področju HaN (Podobnik in dr. 2024b), je zaradi parov slik več modalitet primeren za razvoj in vrednotenje nalog I2I med modalitetami. Poročani rezultati lahko zato služijo kot merilo za prihodnje raziskave o ustvarjanju sCT slik v tem anatomskem območju. Naši rezultati potrjujejo potencial difuzijskih modelov za generacijo visokokakovostnih sCT slik področja HaN. Raziskali smo 2D in 3D pristop generiranja slik, pri čemer je naš 2D model dosegel PSNR 32,1 dB in SSIM 90,3 %, medtem ko je naš 3D model dosegel višji PSNR 33,1 dB in višji SSIM 92,2 %. Medtem ko je pristop z 2D rezinami ustvaril sintetične CT prereze visokih kvalitiet, je 3D pristop bolje ohranil anatomsko kontinuiteto v aksialnih prerezih, kot kažeta koronalni in sagitalni prerez na sliki 1. Čeprav je 3D model dosegel boljše vrednosti metrik, pa je zahteval bistveno več računske moči in daljši čas učenja. Kompromis med anatomsko skladnostjo in računsko učinkovitostjo poudarja potrebo po prilagojenih rešitvah glede na prednostne naloge v klinični praksi. Medtem ko lahko na primer 2D modeli zadostujejo za hitro izdelavo prototipov, so lahko 3D modeli nepogrešljivi za celovite delovne postopke, kot je segmentacija hrbtenice, za katero se zdi, da je anatomska struktura, s katero je imel poleg majhnih organov naš 2D model največ težav. Druga ugotovitev je, da kvantitativne metrike, zlasti MAE, morda ne odražajo v celoti kvalitativnih ugotovitev, prikazanih na sliki 1. Slika 1 namreč jasno kaže, da 2D modeli ustvarjajo nejasne prehode med aksialnimi rezinami in ne dosegajo pričakovane anatomske kontinuitete vzdolž sagitalne in koronalne smeri. Predvidevamo, da bi to omejitev lahko ublažili s pristopom z več rezinami, pri katerem so na primer kot vhodni podatki na voljo tri zaporedne rezine, difuzijski model pa ima nalogo ustvariti le srednjo rezino.

V naši implementaciji v eni učni epohi model ne vidi vseh rezin oziroma koščkov vsake slike iz učne zbirke, temveč se iz vsake slike sproti naključno vzorci 20 rezin ali 2 koščka. Prav tako velja omeniti dolgotrajen postopek inference, saj model potrebuje približno 6 minut, da ustvari vse 2D sCT aksialne rezine, potrebne za sestavo ene zložene 3D sCT slike. Ta čas inference je bil izmerjen z uporabo GPU RTX 2080Ti. Obsežno smo eksperimentirali z zmanjšanjem števila časovnih korakov inference in ugotovili, da se kakovost sintetične slike ni bistveno zmanjšala, če smo namesto 1000 časovnih korakov, uporabljenih za učenje, uporabili 100 časovnih korakov. S tem smo lahko čas inference zmanjšali približno za faktor deset.

Neposredna primerjava z drugimi študijami je zaradi razlik v naborih podatkov in anatomskih regijah težja in ni nujno popolnoma reprezentativna. Zavedajoč se teh omejitev primerjave poročamo, da so Yang in dr. (2020), ki so izvedli prevajanja iz MR v CT na podatkih o možganih, poročali o PSNR 25,2 dB in SSIM 80,1 %, Yan in dr. (2023), ki so izvedli prevajanje iz MR v CT na podatkih HaN, so dosegli PSNR 26,2 dB in SSIM 84,9 %, Graf in dr. (2023) so dosegli PSNR 27,9 dB in SSIM 88,7 % na podatkih o hrbtenici, Wolterink in dr. (2017) pa PSNR 32,3 dB na neparnih možganskih podatkih. Iz zgoraj navedenih rezultatov sorodnih študij lahko sklepamo, da so rezultati v naši študiji boljši od rezultatov v prej omenjenih študijah.

Na podlagi rezultatov na slikah 2 in 3 lahko trdimo, da so segmentacijske maske ustvarjene na podlagi sCT slik, generiranih z našimi modeli, močno podobne segmentacijskim maskam, ki jih isto orodje TotalSegmentator ustvari na podlagi resničnih CT slik. Orodje ima največ težav pri segmentaciji majhnih anatomskih struktur, kot so optični živci in žleza ščitnica, kar je razvidno iz same razpršitve točk na škatlastih diagramih, hkrati pa so ti organi pogosto na robu vidnega polja slik, kar orodju dodatno otežuje nalogo segmentacije. Če primerjamo rezultate metrik 2D in 3D modela, je v povprečju mogoče opaziti višja odstopanja pri 2D modelu za večino anatomskih struktur. V nekaterih primerih so sCT slike od resničnih bolj uporabne za segmentacijske naloge, kar sklepamo ne le iz povprečnih vrednosti metrik, ampak tudi iz primerov, ko uporabljeno orodje TotalSegmentator ni uspelo izvesti segmentacije na resničnih CT slikah, na sCT slikah pa mu je to uspelo.

5 Zaključek

Naučen difuzijski model za prevajanje med MR in CT slikovno modaliteto generira kvalitetne slike, kar dokazujejo visoke vrednosti PSNR in SSIM. Ti obetavni rezultati nakazujejo, da lahko sCT slike ohranijo zadostno anatomsko natančnost za nekatere nadaljnje naloge radioterapije, kot sta orisovanje OAR in načrtovanje obsevanja. Za potrditev te hipoteze so potrebne dodatne obsežne raziskave in empirična potrditev. Metodološki napredki difuzijskih modelov, razpoložljivost podatkovnih zbirk s pari CT in MR slik ter zapletenost paradigme MR-samostojnega načrtovanja zdravljenja v delovnem postopku radioterapije predstavljajo številne priložnosti za nadaljnje raziskave.

Opomba

Delo je odobrila Etična komisija Onkološkega inštituta Ljubljana, Slovenija, pod številko ERID-EK/139, podprla pa ga je Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost Republike Slovenije (ARIS) v okviru projektov P2-0232 in J2-60042.

Viri in literatura

- Cardoso, M. J., Li, W., Brown, R., and et al., “MONAI: an open-source framework for deep learning in healthcare,” *arXiv:2211.02701* (2022).
- Dayarathna, S., Islam, K. T., Uribe, S., Yang, G., Hayat, M., and Chen, Z., “Deep learning based synthesis of MRI, CT and PET: review and analysis,” *Med. Image Anal.* 92, 103046 (2024).
- Graf, R., Schmitt, J., Schlaeger, S., Möller, H. K., Sideri-Lampretsa, V., Sekuboyina, A., Krieg, S. M., Wiestler, B., Menze, B., Rueckert, D., and Kirschke, J. S., “Denoising diffusion-based MRI to CT image translation enables automated spinal segmentation,” *Eur. Radiol. Exp.* 7, 70 (2023).
- Gregoire, V., Guckenberger, M., Haustermans, K., Lagendijk, J. J. W., Ménard, C., Pöötter, R., Slotman, B. J., Tanderup, K., Thorwarth, D., van Herk, M., and Zips, D., “Image guidance in radiation therapy for better cure of cancer,” *Mol. Oncol.* 14, 1470–1491 (2020).
- Ho, J., Jain, A., and Abbeel, P., “Denoising diffusion probabilistic models,” in *[34th Conference on Neural Information Processing Systems - NeurIPS 2020], Advances in Neural Information Processing Systems* 33, 6840–6851, NeurIPS (2020).
- J. Wasserthal, H.-C. Breit, M. T. Meyer, M. Pradella, D. Hinck, A. W. Sauter, T. Heye, D. T. Boll, J. Cyriac, S. Yang, M. Bach in M. Segeroth, “Totalsegmentator: Robust segmentation of 104 anatomic structures in ct images,” *Radiology: Artificial Intelligence*, vol. 5, no. 5, str. e230024, 2023.
- Kazerouni, A., Aghdam, E. K., Heidari, M., Azad, R., Fayyaz, M., Hacihaliloglu, I., and Merhof, D., “Diffusion models in medical imaging: a comprehensive survey,” *Med. Image Anal.* 88, 102846 (2023).
- Liu, L., Ren, Y., Lin, Z., and Zhao, Z., “Pseudo numerical methods for diffusion models on manifolds,” in *[10th International Conference on Learning Representations - ICLR 2022]*, 1–23 (2022).
- Li, Y., Xu, S., Lu, Y., and Qi, Z., “CT synthesis from MRI with an improved multi-scale learning network,” *Front. Phys.* 11, 1088899 (2023).
- McDonald, B. A., Dal Bello, R., Fuller, C. D., and Balermipas, P., “The use of MR-guided radiation therapy for head and neck cancer and recommended reporting guidance,” *Semin. Radiat. Oncol.* 34, 69–83 (2024).
- Podobnik, G., Strojjan, P., Peterlin, P., Ibragimov, B., and Vrtovec, T., “HaN-Seg: The head and neck organ-at-risk CT & MR segmentation dataset,” *Med. Phys.* 50, 1917–1927 (2023).
- Podobnik, G., Ibragimov, B., Peterlin, P., Strojjan, P., and Vrtovec, T., “vOARiability: Interobserver and intermodality variability analysis in oar contouring from head and neck CT and MR images,” *Med. Phys.* 51, 2175–2186 (2024).
- Podobnik, G., Ibragimov, B., Tappeiner, E., Lee, C., Kim, J. S., Mesbah, Z., Modzelweski, R., Ma, Y., Yang, F., Rudecki, M., Wodźniński, M., Peterlin, P., Strojjan, P., and Vrtovec, T., “HaN-Seg: The head and neck organ-at-risk CT and MR segmentation challenge,” *Radiother. Oncol.* 198, 110410 (2024).
- Ronneberger, O., Fischer, P., and Brox, T., “U-Net: convolutional networks for biomedical image segmentation,” in *[18th International Conference on Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015], Lecture Notes in Computer Science* 9351, 234–241, Springer (2015).
- Song, J., Meng, C., and Ermon, S., “Denoising diffusion implicit models,” in *[9th International Conference on Learning Representations - ICLR 2021]*, 1–20 (2021).
- Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P., “Image quality assessment: from error visibility to structural similarity,” *IEEE Trans. Image Process.* 13, 600–612 (2004).

- Wolterink, J. M., Dinkla, A. M., Savenije, M. H. F., Seevinck, P. R., van den Berg, C. A. T., and Išgum, I., “Deep MR to CT synthesis using unpaired data,” in *[2nd International Workshop on Simulation and Synthesis in Medical Imaging - SASHIMI 2017, held in conjunction with MICCAI 2017], Lecture Notes in Computer Science 10557*, 14–23, Springer (2017).
- Yang, H., Sun, J., Carass, A., Zhao, C., Lee, J., Prince, J. L., and Xu, Z., “Unsupervised MR-to-CT synthesis using structure-constrained CycleGAN,” *IEEE Trans. Med. Imaging* 39, 4249–4261 (2020).
- Yang, H., Sun, J., Carass, A., Zhao, C., Lee, J., Prince, J. L., and Xu, Z., “Unsupervised MR-to-CT synthesis using structure-constrained CycleGAN,” *IEEE Trans. Med. Imaging* 39, 4249–4261 (2020).

AIM@VET-INSPIRED UNIVERSITY LEVEL EDUCATION STRATEGIES FOR TEACHING COMPUTER VISION AND BIOMETRICS

ŽIGA EMERŠIČ,¹ GREGOR HRASTNIK,² NATAŠA MEH PEER,²
PETER PEER¹

¹ University of Ljubljana, Faculty of Computer and Information Science, Ljubljana, Slovenia

ziga.emersic@fri.uni-lj.si, peter.peer@fri.uni-lj.si

² School center Velenje, Velenje, Slovenia

gregor.hrastnik@scv.si, natasa.mehpeer@scv.si

Computer vision and biometrics are increasingly important in many AI-driven applications, yet teaching these fields poses challenges in balancing theory and hands-on practice. This paper presents a structured approach implemented for the technical skills course at the Faculty of Computer and Information Science, University of Ljubljana, designed for Computer Science students. The course integrates guided Jupyter Notebook exercises while allowing students to complete coding tasks while leaning on AI assistance. In-person presentations and discussions reinforce understanding by requiring students to explain their implementations and problem-solving strategies. The 15-week curriculum progresses from basic image processing to deep learning-based biometric recognition. Teaching materials are derived from the AIM@VET EU project, which focuses on adapting AI education to labor market needs, but adapted here for university students. We hope that AI-assisted, structured coding exercises combined with interactive discussions will enhance engagement and comprehension, better preparing students for a variety of applications in computer vision and biometrics.

DOI

[https://doi.org/
10.18690/um.feri.2.2025.4](https://doi.org/10.18690/um.feri.2.2025.4)

ISBN

978-961-286-960-1

Keywords:

computer vision,
biometrics,
artificial intelligence
education,
ai-assisted learning,
deep learning,
AIM@VET



University of Maribor Press

DOI
[https://doi.org/
10.18690/um.feri.2.2025.4](https://doi.org/10.18690/um.feri.2.2025.4)

ISBN
978-961-286-960-1

AIM@VET IZOBRAŽEVALNI MATERIALI ZA UNIVERZITETNO POUČEVANJE RAČUNALNIŠKEGA VIDA IN BIOMETRIJE

ŽIGA EMERŠIČ¹ GREGOR HRASTNIK,² NATAŠA MEH PEER,²
PETER PEER¹

¹ Univerza v Ljubljani, Fakulteta za računalništvo in informatiko, Ljubljana, Slovenija
ziga.emersic@fri.uni-lj.si, peter.peer@fri.uni-lj.si

² Šolski center Velenje, Velenje, Slovenija
gregor.hrastnik@scv.si, natasa.mehpeer@scv.si

Ključne besede:

računalniški vid,
biometrija,
izobraževanje o umetni
inteligenci,
učenje s podporo umetne
inteligence,
globoko učenje,
AIM@VET

Računalniški vid in biometrija postajata vse pomembnejša na različnih področjih umetne inteligence, vendar njuno poučevanje predstavlja izziv pri uravnoteženju teorije in praktičnega dela. Ta članek predstavlja strukturiran pristop, uporabljen za predmet tehničnih veščin na Fakulteti za računalništvo in informatiko Univerze v Ljubljani, namenjen študentom računalništva. Predmet vključuje vodene vaje v Jupyter Notebookih in študentom omogoča reševanje programskih nalog s pomočjo raznih inteligentnih orodij. Razumevanje snovi je dodatno okrepljeno s predstavitvami in razpravami, kjer študenti predstavijo svoje rešitve in reševanje problemov. Predmet v 15 tednih napreduje od osnovnega procesiranja slik do biometričnega prepoznavanja z globokim učenjem. Učni materiali izhajajo iz evropskega projekta AIM@VET, ki se osredotoča na prilagajanje izobraževanja o umetni inteligenci potrebam trga dela, tu pa so prilagojeni za univerzitetne študente. Upamo, da naloge podprte z umetno inteligenco in interaktivne razprave izboljšujejo angažiranost in razumevanje ter študente bolje pripravljajo na različna področja računalniškega vida in biometrije.



Univerzitetna založba
Univerze v Mariboru

1 Introduction

Computer vision and biometrics are rapidly advancing fields with widespread applications in security, healthcare, autonomous systems, and human-computer interaction. As these technologies continue to shape various industries, it is essential for university students to gain both a strong theoretical foundation and practical experience. However, teaching these subjects at the university level presents several challenges. Many core algorithms rely on complex mathematical concepts and deep learning techniques, which can be difficult for students to grasp without proper guidance and hands-on experience. Furthermore, computer vision and biometrics demand significant computational resources, often requiring high-performance GPUs and large datasets to train and evaluate models effectively. This creates additional barriers for students who may not have access to advanced hardware. Additionally, the interdisciplinary nature of these fields means that students must also understand their connections to artificial intelligence, cybersecurity, and ethical considerations, making it necessary to design a curriculum that balances multiple disciplines (Jiang, 2023; Wu, 2023).

To address these challenges, we have developed educational materials tailored specifically for university-level Computer Science students. The materials are based on parts of the ongoing EU project AIM@VET (Artificial Intelligence Modules for Vocational Education and Training) (AIM@VET 2025), which focuses on developing learning modules to adapt Vocational Education and Training to the needs of the labor market, with a particular emphasis on AI. The project involves six partners from Spain, Portugal, and Slovenia. In the Slovenian branch, the University of Ljubljana (UL) and School Center Velenje (SCV) serve as contributing partners, focusing on the computer vision aspect of AI. The topics covered include capturing and curating unbiased data, detection and segmentation, and tracking and recognition. This closely aligns with the needs of the developed university-level educational materials, albeit at a more advanced level. Furthermore, during the development of the AIM@VET materials, some parts were rejected or modified due to their excessive complexity (Emeršič, 2023; Kirn 2024; Emeršič, 2024; Emrešič, 2025a; Emeršič 2025b); these were instead incorporated into the proposed university-level materials.

The teaching materials will be used in the course "Technical Skills 2" at the Faculty of Computer and Information Science, University of Ljubljana. The course will introduce students to fundamental concepts of computer vision, with a strong focus on biometrics, and will incorporate deep learning techniques alongside extensive hands-on activities. The curriculum is designed to bridge the gap between theory and practice by integrating modern frameworks such as OpenCV and PyTorch, enabling students to build real-world biometric systems, including detection, segmentation and recognition models. Given the increasing demand for these skills, the course emphasizes project-based learning, encouraging students to engage with real datasets, implementing machine learning models, and optimize biometric authentication methods. The Technical Skills course will, hopefully, not only prepare students for careers in AI-driven fields but also ensure they develop problem-solving abilities applicable across multiple domains. This paper summarized the educational approaches implemented in the course and the solutions adopted to enhance the learning experience for students specializing in computer vision and biometrics.

2 Educational Approaches

2.1 Traditional vs. Modern Teaching Methods

Teaching computer vision and biometrics requires a combination of traditional theoretical instruction and modern, hands-on learning techniques. Traditional methods, such as textbook learning and lectures, provide a necessary foundation for understanding the mathematical principles and algorithms underlying computer vision. These include concepts like convolution, feature extraction, and statistical learning, which are essential for understanding advanced AI techniques. However, when presented in isolation, these methods can be abstract and disengaging for students, especially in a field that relies heavily on practical applications (Ashwin, 2023; Jeon, 2023; Abdrakhmanov, 2024).

To bridge this gap, modern teaching approaches incorporate interactive coding exercises, real-world applications, and AI-driven tools to enhance engagement. Instead of focusing solely on theoretical derivations, students actively experiment with image processing, biometric recognition, and deep learning models in programming environments. The integration of frameworks like OpenCV, TensorFlow, and PyTorch allows students to visualize and manipulate real-world

data, improving their understanding of key concepts. This hands-on approach ensures that students not only grasp theoretical principles but also develop skills applicable in real-world industry scenarios.

2.2 Hands-On Learning and AI Assistance

To ensure practical engagement, the course follows a structured hands-on learning approach, where students work with pre-prepared Jupyter scripts that contain partially implemented code. Students are required to fill in missing parts, apply learned concepts, and experiment with different methods to complete tasks. This structured guidance helps them progressively build confidence in implementing complex algorithms while still being encouraged to think critically and experiment with solutions rather than just following instructions.

Another feature of this course is the integration of modern AI tools, such as ChatGPT and other large language models (LLMs), into the learning process. Students are encouraged to use these tools to troubleshoot code, generate explanations, and explore alternative solutions. However, reliance on LLM-generated content is not the final step—students must present and discuss their work in person, explaining their approach, reasoning, and any issues encountered. This ensures they fully understand the concepts rather than blindly following AI-generated solutions.

2.3 Assessment Strategies

To ensure that students acquire a balanced mix of theoretical understanding and practical skills, multiple assessment methods will be employed:

- Exams – Evaluate foundational knowledge, including mathematical concepts and algorithmic principles.
- Practical Assignments – Small coding exercises where students complete Jupyter Notebook tasks by implementing missing functionality.
- Final Projects – Require students to apply their knowledge to develop fully functional biometric systems.
- In-Person Presentations & Discussions – Students must present their solutions, explain their decision-making process, and discuss challenges they

encountered. This encourages critical thinking and ensures they understand the material beyond simply writing functional code.

These assessments ensure that students develop both conceptual understanding and hands-on expertise in computer vision and biometrics, while also improving their ability to critically evaluate AI-generated outputs.

3 Challenges and Solutions

3.1 Mathematical and Algorithmic Complexity

One of the primary challenges in teaching computer vision and biometrics is the complexity of the underlying mathematics. Many students struggle with concepts such as convolution, eigenvalues, and optimization techniques, which are fundamental to image processing and deep learning. To address this, the course incorporates interactive visualizations and step-by-step coding exercises that allow students to see how these mathematical concepts translate into real-world applications. Additionally, simplifying explanations before introducing complex equations helps students gradually build a solid understanding.

3.2 Hardware & Infrastructure Needs

Training deep learning models for computer vision and biometric analysis typically requires high-performance GPUs and large datasets, which can be a challenge for students working on personal machines. To mitigate this, the course provides access to cloud-based computing platforms such as Google Colab and AWS, allowing students to run deep learning experiments without requiring expensive hardware. Additionally, the use of pre-trained models helps reduce computational costs while still allowing students to experiment with biometric recognition tasks.

3.3 Interdisciplinary Nature

Computer vision and biometrics intersect with multiple disciplines, including AI, cybersecurity, ethics, and statistics, which can make it difficult for students to grasp the full scope of the field. The course addresses this by integrating interdisciplinary modules that encourage students to explore real-world applications. For example,

students will study biometric authentication systems from both a technical and ethical perspective, learning about issues such as bias in biometric algorithms and security vulnerabilities in facial recognition systems.

3.4 Industry Relevance & Employability

Employers in AI-driven industries prioritize practical implementation skills over theoretical knowledge alone. Many university courses fail to equip students with the hands-on experience necessary for careers in computer vision and biometrics. To bridge this gap, the course includes industry-relevant projects, guest lectures from experts in the field, and opportunities for students to contribute to open-source projects. Additionally, by requiring students to present and defend their work in discussions, the course strengthens their ability to explain technical concepts, a skill that is highly valued in both industry and academia.

4 The Proposed Teaching Materials

The proposed course, "Technical Skills 2", at the Faculty of Computer and Information Science, University of Ljubljana, follows a structured 15-week plan, covering key topics in computer vision and biometrics. The teaching materials have been developed to align with industry needs while ensuring that students gain both foundational and advanced knowledge in the field. The course follows a progressive structure, where students begin with basic image processing and gradually advance to deep learning-based biometric recognition. An example of a Jupyter script with solutions is shown in Figure 1.

Course Breakdown:

- Weeks 1–2: Introduction to computer vision fundamentals, image manipulation, and ethics in AI. Students learn how to import, process, and enhance images while addressing issues like bias in biometric datasets.
- Weeks 3–5: Students explore biometric modalities (face, fingerprints, gait, voice, etc.), implement object detection using classical methods, and transition to deep learning-based detection models like YOLO and Fast R-CNN.

- Weeks 6–7: Introduction to semantic segmentation, including U-Net and Mask R-CNN, applied to biometric data.
- Weeks 8–10: Object tracking techniques and advanced human recognition methods such as gait and action recognition.
- Weeks 11–12: Feature extraction and pattern analysis methods for biometric authentication.
- Weeks 13–14: Deep learning approaches for biometric recognition using ResNet, EfficientNet, and ViT.
- Week 15: Integration of individual components into a complete biometric recognition pipeline, including real-world use cases.

```

Crate our own image Transformer using Pytorch

# Help yourself with the provided code above and create a transformation that adds a random noise to the image
# We can imagine it like like a dot or speck on our numbers, when we take an image of a number
# The image we would take would not be perfect, it would have some noise (salt, pepper, ...) so it

class RandomSaltPepperTransform:
    def __init__(self, num_specks=9, speck_size=1):
        ...

    def __call__(self, image):
        # Convert the image to a tensor if it's not already
        # Get the dimensions of the image
        # Add random specks to the image
        # Randomly choose the position of the speck
        # Add the speck to the image
        return image

# Define transformations
transform = Transform.Compose([
    transforms.Resize((28, 28)), # resize the image to 28x28 pixels
    transforms.Grayscale(num_output_channels=1), # convert the image to grayscale (1 channel)
    RandomSaltPepperTransform(),
])

dataset = ImageFolder(root_path, # root directory of the dataset
                    transform=transform) # specify the transformations to be applied on the images
dataloader = DataLoader(dataset, # dataset to load
                        batch_size=9, # how many samples per batch to load
                        shuffle=True) # set to True to have the data reshuffled at every epoch
                        # we shuffle the data to avoid overfitting

# Get a batch of training data
dataset_iter = iter(dataloader)
images, labels = next(dataset_iter)

# Show images with labels
# torchvision.utils.make_grid(images, labels)

# Sample image tensor
image_tensor = images[0]
print(f"Image Tensor Shape: {image_tensor.shape}")
print(f"Dimensions: (Channels, Height, Width)")

```

```

Crate our own image Transformer using Pytorch

# Help yourself with the provided code above and create a transformation that adds a random noise to the image
# We can imagine it like like a dot or speck on our numbers, when we take an image of a number
# The image we would take would not be perfect, it would have some noise (salt, pepper, ...) so it

class RandomSaltPepperTransform:
    def __init__(self, num_specks=9, speck_size=1):
        self.num_specks = num_specks # Number of specks to add
        self.speck_size = speck_size # Size of each speck

    def __call__(self, image):
        # Convert the image to a tensor if it's not already
        image = transforms.ToTensor()(image)
        # Get the dimensions of the image
        # Get the width, height, and image shape
        # Add random specks to the image
        for _ in range(self.num_specks):
            # Randomly choose the position of the speck
            x = random.randint(0, width - 1)
            y = random.randint(0, height - 1)
            # Add the speck to the image
            image[x, y] += torch.FloatTensor([1.0])

        return image

# Define transformations
transform = Transform.Compose([
    transforms.Resize((28, 28)), # resize the image to 28x28 pixels
    transforms.Grayscale(num_output_channels=1), # convert the image to grayscale (1 channel)
    RandomSaltPepperTransform(),
])

dataset = ImageFolder(root_path, # root directory of the dataset
                    transform=transform) # specify the transformations to be applied on the images
dataloader = DataLoader(dataset, # dataset to load
                        batch_size=9, # how many samples per batch to load
                        shuffle=True) # set to True to have the data reshuffled at every epoch
                        # we shuffle the data to avoid overfitting

# Get a batch of training data
dataset_iter = iter(dataloader)
images, labels = next(dataset_iter)

# Show images with labels
torchvision.utils.make_grid(images, labels)

# Sample image tensor
image_tensor = images[0]
print(f"Image Tensor Shape: {image_tensor.shape}")
print(f"Dimensions: (Channels, Height, Width)")

```

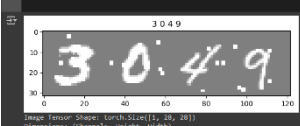


Image Tensor Shape: torch.Size([1, 28, 28])
Dimensions: (Channels, Height, Width)

Figure 1: An example of teaching material in the form of a Jupyter script. On the left, a framework for completion; on the right, a display of the completed solution.

Source: Own work.

The educational materials draw upon the AIM@VET EU project, which focuses on adapting AI education to meet labor market demands. While AIM@VET primarily targets vocational education, some more advanced content that was deemed too complex for vocational training has been adapted for university-level instruction

(Emeršič, 2023; Kirn, 2024). These materials include advanced deep learning models, biometric security case studies, and practical applications in AI ethics.

By structuring the course around hands-on activities and real-world applications, students gain the skills necessary to develop industry-grade biometric recognition systems. The combination of theoretical lectures, structured coding exercises, LLM-assisted problem-solving, and in-person discussions ensures that graduates are well-equipped for careers in AI, cybersecurity, and biometric research.

5 Conclusion

This paper demonstrates a structured, hands-on approach to teaching computer vision and biometrics, combined with AI-assisted learning and interactive discussions. Preliminary findings during AIM@VET project, show that this enhances student engagement and comprehension. Materials adapted and implemented for the course "Technical Skills 2" at the Faculty of Computer and Information Science, University of Ljubljana, the methodology integrates guided Jupyter Notebook exercises, biometric datasets, and modern LLMs, such as OpenAI's, Meta's and DeepSeek's derivatives, enabling students to complete coding tasks while critically evaluating AI-generated solutions. The requirement for in-person presentations and discussions further reinforces understanding by encouraging students to articulate their reasoning and problem-solving strategies.

The 15-week curriculum, covering topics from basic image processing to deep learning-based biometric recognition, provides students with both theoretical foundations and industry-relevant practical skills. Challenges such as mathematical complexity and computational resource limitations are addressed through cloud-based computing environments and interdisciplinary project design. Teaching materials, partially derived from the AIM@VET EU project, align with industry needs and ensure that students gain practical experience applicable to real-world biometric and AI-driven applications.

Findings during AIM@VET suggest that combining structured coding exercises, AI tools, and interactive discussions fosters a deeper understanding of computer vision and biometrics while preparing students for their careers in AI. Future work could explore further AI-driven enhancements in education, such as automated feedback

systems during learning, adaptive learning pathways, and deeper industry collaborations, to continue improving the effectiveness of AI education.

References

- Abdrakhmanov, R. et al. (2024). Applying computer vision and machine learning techniques in education. *International Journal of Advanced Computer Science and Applications*, 15(1). https://thesai.org/Downloads/Volume15No1/Paper_82-Appling_Computer_Vision_and_Machine_Learning_Techniques.pdf
- AIM@VET. AIM@VET Guidelines. Pridobljeno iz: <https://aim4vet.udc.es/backgroud/> (2025).
- Ashwin, M. (2023). A study on various applications of computer vision for teaching learning in classroom. *IEEE Conference Publication*. <https://ieeexplore.ieee.org/document/10009136/>
- Emeršič, Ž., Hrastnik, G., Meh Peer, N., Peer, P. (2023). »Adapting VET Education to Labor Market Needs with Focus on Artificial Intelligence and Computer Vision«, ROSUS 2023
- Emeršič, Ž., Hrastnik, G., Meh Peer, N., Kirn, V. L., Justin, A., Videnović, J., Markičević, L., & Peer, P. (2025a). Educational computer vision materials for classification and tracking of objects. In V. Julian (Ed.), *Intelligent data engineering and automated learning - IDEAL 2024. Part 2: 25th International Conference, Valencia, Spain, November 20–22, 2024: Proceedings (Lecture Notes in Computer Science, Vol. 15347, pp. 502–507)*. Springer. https://doi.org/10.1007/978-3-031-77738-7_42
- Emeršič, Ž., Hrastnik, G., Meh Peer, N., Kirn, V. L., Justin, A., Videnović, J., Markičević, L., Peer, P. (2025b). Simplification of image segmentation and object detection teaching materials. In V. Julian (Ed.), *Intelligent data engineering and automated learning - IDEAL 2024. Part 2: 25th International Conference, Valencia, Spain, November 20–22, 2024: Proceedings (Lecture Notes in Computer Science, Vol. 15347, pp. 495–501)*. Springer. https://doi.org/10.1007/978-3-031-77738-7_41
- Emeršič, Ž., Peer, P., Hrastnik, G., Meh Peer, N., Bey, J. M., Meizoso-García, M., Silva, A. P., Domingues, C., Abreu, C., Costa, A., Durães, D., Novais, P., Renda, C., & Prieto, A. (2024). Integrating AI into VET: Insights from AIM@VET's first training activity. In S. Kadry (Ed.), *Artificial intelligence and education: Shaping the future of learning (Vol. 28, pp. 95–120)*. IntechOpen. <https://doi.org/10.5772/intechopen.1004949>
- Jeon S. et al. (2023). A staged framework for computer vision education: Integrating AI. *Applied Sciences*, 14(21), 9792. <https://www.mdpi.com/2076-3417/14/21/9792>
- Jiang, H., Fu, W. (2023). Computer vision recognition in the teaching classroom: A review. *EAI Endorsed Transactions on AI and Robotics*. <https://publications.eai.eu/index.php/airo/article/view/4079>
- Kirn, V. L., Emeršič, Ž., Hrastnik, G., Meh Peer, N., Peer, P. (2024). »Introductory Computer Vision Teaching Materials for VET Education«, ROSUS 2024
- Wu, T.T., et al. (2023). Leveraging computer vision for adaptive learning in STEM education. (2023). *International Journal of Educational Technology in Higher Education*. <https://educationaltechnologyjournal.springeropen.com/articles/10.1186/s41239-023-00422-5>

SEGMENTACIJA REGIJ ZANIMANJA V HISTOPATOLOŠKIH POSNETKIH KOLOREKTALNIH POLIPOV

MARTIN ŠAVC, BOŽIDAR POTOČNIK

Univerza v Mariboru, Fakulteta za elektrotehniko računalništvo in informatiko, Maribor,
Slovenija

martin.savc@um.si, bozidar.potocnik@um.si

Histopatološki posnetki pogosto vsebujejo veliko diagnostično nepomembnih motečih informacij. Patolog se mora pri svojem delu osredotočiti na specifične regije, kjer opazuje tako podrobnosti kot obliko in število večjih celičnih struktur. V tem delu predstavimo dva poskusa označevanja regij zanimanja in učenje modelov segmentacije za avtomatsko detektiranje teh regij. Prvi pristop je bil t. i. grobo označevanje, ki je za označevalca manj naporen in časovno učinkovitejši. V tem eksperimentu je bilo označenih 123 slik. Izkazalo se je, da je bil model segmentacije, naučen s temi podatki, bolj natančen kot pa same pripravljene oznake. Drugi pristop je bil t. i. podrobno označevanje, ki je za označevalca bistveno bolj naporen. S tem označevanjem je bilo označenih samo 10 slik. Model naučen s temi podatki je, kljub izredno majhnemu naboru učnih podatkov, ob vizualnem pregledu bolje segmentiral regije zanimanja kot pa model naučen z grobimi oznakami.

DOI

[https://doi.org/
10.18690/um.feri.2.2025.5](https://doi.org/10.18690/um.feri.2.2025.5)

ISBN

978-961-286-960-1

Ključne besede:

histopatološki posnetki,
segmentacija regij
zanimanja,
globoko učenje,
kolorektalni polipi,
obdelava medicinskih slik



Univerzitetna založba
Univerze v Mariboru

DOI
[https://doi.org/
10.18690/um.feri.2.2025.5](https://doi.org/10.18690/um.feri.2.2025.5)

ISBN
978-961-286-960-1

Keywords:

histopathological images,
region of interest
segmentation,
deep learning,
colorectal polyps,
medical image processing

REGION OF INTEREST SEGMENTATION IN HISTOPATHOLOGICAL IMAGES OF COLORECTAL POLYPS

MARTIN ŠAVC, BOŽIDAR POTOČNIK

University of Maribor, Faculty of Electrical Engineering and Computer Science,
Maribor, Slovenia
martin.savc@um.si, bozidar.potocnik@um.si

Histopathological images often contain a lot of diagnostically irrelevant, distracting information. The pathologist needs to focus on specific regions where he can observe details as well as the shape and number of larger cellular structures. In this paper, we present two approaches to labelling regions of interest and learning segmentation models for automatic detection of these regions. The first approach was so-called coarse labelling, which is less laborious and more time-efficient for the labeller. In this experiment, 123 images were labelled. It turned out that the segmentation model trained on this data was more accurate than the labels themselves. The second approach was the so-called fine labelling, which is much more time-consuming for the labeller. Only 10 images were labelled using this method. Despite the extremely small training data set, the model trained with this data segmented the regions of interest better than the model trained with coarse labels.



1 Uvod

Analiza kolorektalnih vzorcev je pomembno orodje za zgodnjo detekcijo in zdravljenje raka na debelem črevesu. Tipično visoko izobražen specialist (patolog) izvede histopatološki pregled, kjer ustrezno pripravi vzorec tkiva ter ga vizualno analizira z mikroskopom. Vizualna analiza je seveda časovno zamudna, saj strokovnjak izbira pomembne regije najprej pri manjši povečavi tkiva, kandidatne regije pa nato podrobneje analizira pri večjih povečavah. Podrobnosti so ključnega pomena za diagnosticiranje. Cilj je torej popolna avtomatizacija delovnega procesa, ki bi razbremenila strokovnjake ročnega preiskovanja takšnih posnetkov. V več raziskovalnih delih (Madabhushi et al. 2016, Song et al. 2020, Šavc et al. 2023) je bilo demonstrirano, da je postopek analiziranja posnetkov možno v veliki meri avtomatizirati s pomočjo ustrezne programske opreme. Nekaj zadnjih uspešnejših rešitev temelji na globokih nevronskih mrežah.

V naši raziskovalni skupini raziskujemo na tem problemskem področju že kar nekaj zadnjih let. Pri razvoju svojih rešitev uporabljamo podatkovno zbirko UniToPatho. Ta zbirka z označenimi podatki je prvenstveno namenjena razvoju računskih metod za razvrščanje vzorcev tkiv. V zbirki najdemo vzorce iz 6 razredov tkiv, in sicer iz razreda hiperplastično (HP), normalno (NORM), tubularni adenom nizke stopnje (TA.LG) in visoke stopnje (TA.HG), ter tubularno vilozni adenom nizke stopnje (TVA.LG) in visoke stopnje (TVA.HG). Podrobnosti o zbirki UniToPatho ter o diagnostičnem pomenu posameznih vrst tkiva so zbrane v (Barbano et al. 2021). Naj omenimo le še to, da ta zbirka sestoji iz označenih 7000 μm slik (velikosti 15855x15855 pikslov) in označenih 800 μm slik (velikosti 1812x1812 pikslov), ki so bile dobljene z razrezom originalnih mikroskopskih posnetkov (angl. Whole-Slide Images, WSI) na patologu pomembne poddele. Po posvetovanju z lokalnim strokovnjakom za analiziranje histopatoloških slik smo se seznanili, da specialisti sprejemajo odločitev o vrsti tkiva na osnovi slik, zajetih pri višji ločljivosti, saj le na takšen način dobijo ustrezen vpogled v tkivo. V naših raziskavah smo se zato osredotočili na 867 vzorcev tkiva velikosti 7000 μm iz zbirke UniToPatho.

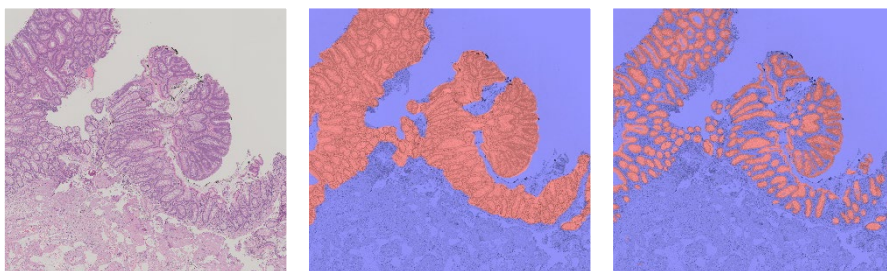
Posamezno 7000 μm tkivo vsebuje ob za razvrščanje koristnih informacijah (tj. o celicah), tudi mnogo drugih nepomembnih informacij kot, na primer, tekočino in fibrovaskularno jedro (glej sliko 1, levo). Pri svojem raziskovalnem delu uporabljamo globoke nevronske mreže. Za njih je značilno, da so se zmožne bolj ali manj uspešno

priučiti značilnosti iz učnih podatkov, s čimer lahko rešimo določeno nalogo. Problem pri obdelovanju slik 7000 μm tkiva z globokimi nevronskimi mrežami pa je v izjemni količini podatkov, ki je današnji računalniški sistemi niso zmožni obdelati naenkrat. Pri slikah takšnih velikosti namreč trčimo na omejitev strojne opreme, tako glede pomnilniških (omejena količina pomnilnika na enoti GPU) kot tudi glede računskih kapacitet (omejene računske zmogljivosti). Ponuja se nam sicer rešitev v obliki pomanjševanja slik, vendar pri analiziranju histopatoloških slik to ni najprimernejše, saj so za diagnosticiranje izrednega pomena tudi podrobnosti. Na našem domenskem področju je najbolj naraven pristop torej ta, da iz ekstremno velikih slik z ustreznimi postopki odstranimo čim več nekoristne informacije, in zgolj preostalo informacijo posredujemo nevronske mreži v postopku učenja.

V tej raziskavi predstavljamo postopek detektiranja regij zanimanja (angl. Region of Interest, ROI) v histopatoloških posnetkih kolorektalnih polipov. Naša rešitev temelji na uveljavljeni U-arhitekturi nevronskih mrež, kjer smo za kodirnik uporabili uveljavljeno konvolucijsko nevronske mrežo VGG11, dekodirnik pa preprosto sestavili z zaporedjem transponiranih konvolucijskih slojev. Globoko nevronske mreže smo učili nadzorovano s pomočjo lastno označenih podatkov iz zbirke UniToPatho.

2 Regije zanimanja

Strokovnjak-patolog se pri analiziranju histopatoloških slik kolorektalnih polipov fokusira na posamezna področja slike, na t. i. področja zanimanja (ROI). Odločitev o tem, s kakšno vrsto tkiva ima opravka, posledično pa tudi o tveganjih za pacienta, sprejme tako na osnovi globalnega pogleda na področje zanimanja (barva, globalna struktura itn.), kakor tudi s podrobnim analiziranjem določenih poddelov regij zanimanja (spremembe v barvi, teksturi in strukturi). Na sliki 1 (levo) vidimo primer originalne 7000 μm histopatološke slike iz zbirke UniToPatho. Na sredinski sliki so prikazane grobo označene regije zanimanja, položene na originalno mikroskopsko sliko, na desni sliki pa vidimo še primer podrobno označenih regij zanimanja.



Slika 1: Primer histopatološkega posnetka iz zbirke UniToPatho (levo), grobo označene (sredina) in podrobno označene (desno) regije zanimanja. Rdeča barva predstavlja regije zanimanja.

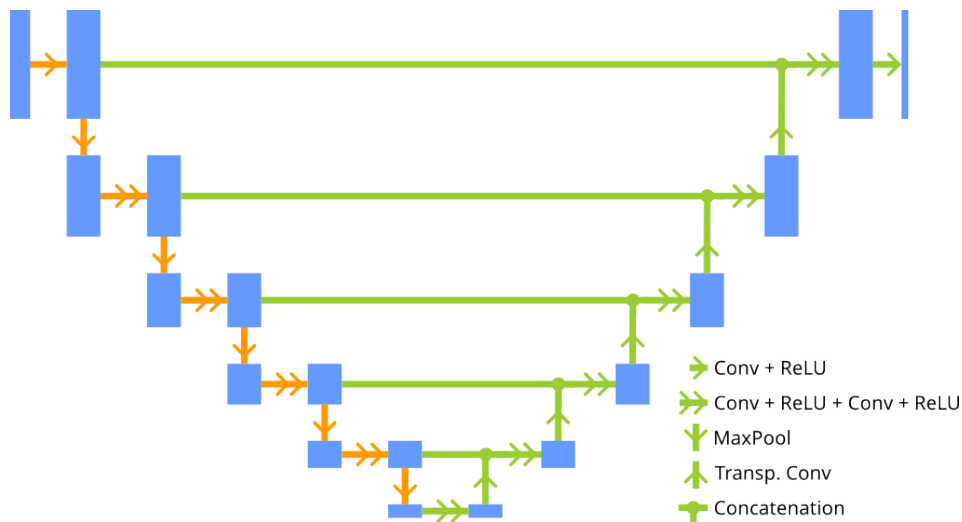
Vir: lasten

3 Segmentacijski model

Problem segmentiranja regij zanimanja v histopatoloških posnetkih kolorektalnih polipov smo v tej raziskavi reševali s pomočjo popolnoma konvolucijske nevronske mreže (angl. fully convolutional neural network). Našo rešitev temeljimo na arhitekturi VGG11 (Simonyan et al. 2014). Od le-te smo ohranili konvolucijski del, glavo za klasifikacijo pa smo zavrgli. Ostalo nam je 20 slojev, iz katerih smo pridobili 6 ločljivostnih nivojev. Te smo nato dopolnili v t. i. kodirnik-dekodirnik arhitekturo namenjeno segmentiranju 2D slik, pri čemer smo se zgledovali po arhitekturi nevronske mreže U-Net (Ronneberger et al. 2015). Arhitektura naše nevronske mreže je prikazana na sliki 2.

Za povečanje dimenzionalnosti so bile v delu dekodirnika uporabljene transponirane konvolucije. Slike značilnic iz spodnjih nivojev so bile konkatenerane po kanalih s slikami značilnic kodirnikov na istem nivoju in obdelane z dvema dodatnima konvolucijama.

Na vrhu naše mreže smo dodali še konvolucijski sloj, kjer smo uporabili filter z jedrom velikosti 1x1 in sigmoidno aktivacijsko funkcijo, kar tvori segmentirano sliko.



Slika 2: Arhitektura uporabljene nevronske mreže. Oranžni operatorji so del osnovne VGG11. Zeleni operatorji so bili dodani za dopolnitev v kodirnik-dekodirnik arhitekturo.

Vir: lasten

4 Učenje

4.1 Učni podatki

V tej raziskavi smo izvedli dva eksperimenta: i) eksperiment s pomočjo grobo označenih regij zanimanja ter ii) eksperiment s podrobno/natančno označenimi regijami zanimanja. Naj poudarimo, da smo v obeh eksperimentih uporabili enak segmentacijski model. V nadaljevanju opisujemo podrobnosti obeh izvedenih eksperimentov.

V prvem eksperimentu smo najprej grobo označili regije zanimanja. Ročno smo označili 123 naključno izbranih slik iz zbirke UniToPatho, in sicer 19 slik iz razreda HP, 17 slik iz razreda NORM, 55 slik iz razreda TA.LG, 9 slik iz razreda TA.HG, 19 slik iz razreda TVA.LG ter 4 slike iz razreda TVA.HG. Naj poudarimo, da smo slike naključno izbirali iz učno-validacijskih množic zbirke UniToPatho. Označene slike smo nato za ta eksperiment naključno razdelili v 98 slik za učni in 25 slik za testni nabor.

Podobno smo postopali v drugem eksperimentu, kjer smo na 10 naključno izbranih slikah iz zbirke UniToPatho podrobno označili regije zanimanja. Uporabili smo naslednji nabor slik: 2 slike iz razreda NORM, 2 slike iz razreda TA.LG, 4 slike iz razreda TVA.LG ter 2 slike iz razreda TVA.HG. Naj omenimo, da označene slike sicer niso iz vseh šestih razredov tkiv, a ker rešujemo segmentacijski problem to ne predstavlja bistvene omejitve našega pristopa. Podrobno označene slike smo na koncu naključno razdelili v 8 slik za učni in 2 slike za testni nabor.

Na kratko opišimo še postopek predobdelave slik. Slike v ločljivosti 7000 um smo še pred učenjem prevzorčili na 1/5 originalne velikosti. S preizkušanjem smo ugotovili, da je to še sprejemljiv kompromis med natančnostjo segmentacije in omejitvami strojne opreme. Iz prevzorčenih slik se pri učenju naključno izrežejo kvadratni vzorci velikosti 512x512 pikslov. V fazi bogatenja podatkov (angl. augmentation) se ti vzorci nato naključno vertikalno in horizontalno zrcalijo, prevzorčijo za naključni skalirni faktor med 0,8 in 1,2 ter rotirajo za naključni kot med -90 in 90 stopinj. Vrednosti vzorcev še normaliziramo s povprečjem in standardnim odklonom, izračunanim za mrežo VGG11 na zbirki ImageNet (Deng et al. 2009). Na koncu z verjetnostjo 0,5 spremenimo svetlost, kontrast, saturacijo in odtonek vzorcev za naključni faktor med 0,8 in 1,2.

4.2 Nastavitve učenja

Na kratko predstavimo pomembnejše hiperparametre, uporabljene med učenjem. Uteži VGG11 operatorjev so bile inicializirane iz modela učenega na zbirki ImageNet. Uteži dodanih slojev pa smo inicializirali naključno. Za funkcijo izgube smo uporabili binarno križno entropijo.

Globoko nevronska mrežo smo učili z algoritmom Adam (Kingma et al. 2015) ter s stopnjo učenja 10^{-4} , uporabili pa smo pakete učenja velikosti 8. Shranili smo izključno zadnji model mreže med učenjem.

4 Rezultati

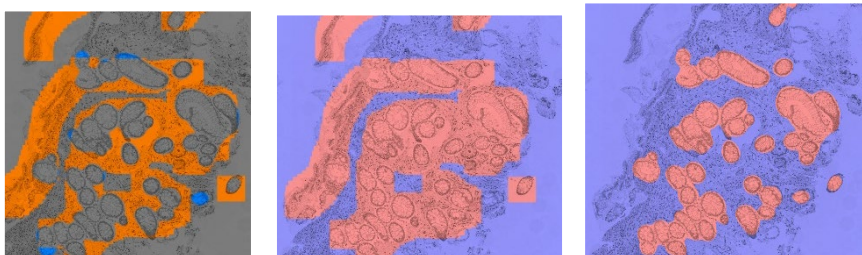
V nadaljevanju bomo najprej predstavili dobljene kvantitativne rezultate obeh izvedenih eksperimentov. Prvem eksperimentu smo mrežo učili z grobimi oznakami. Učenje smo ustavili po 300 epohah (oz. $300 \times 98 = 29.400$ korakov). Tako naučena

mreža je po 300 epohah dosegla na testni množici natančnost (ang. precision) 0,96 in priklic (ang. recall) 0,86.

V drugem eksperimentu pa smo mrežo učili s pomočjo podrobnih označb regij zanimanja. V tem primeru je naučena mreža po 2500 epohah (oz. $2500 \times 8 = 20.000$ korakov) dosegla na testni množici natančnost 0,99 in priklic 0,87.

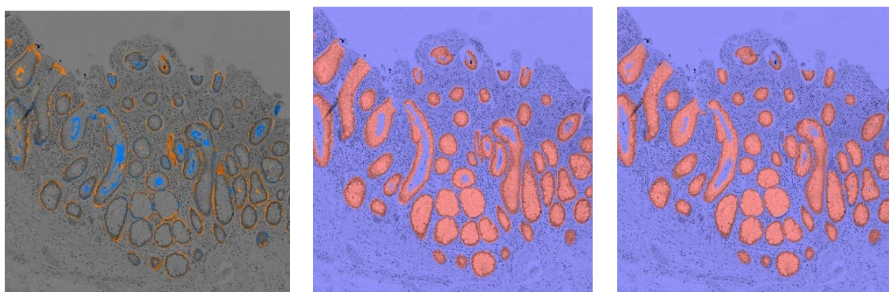
Oglejmo si še nekaj kvalitativnih rezultatov. Ker se je postopek učenja v obeh eksperimentih izkazal z visoko natančnostjo in priklicem za uspešnega, bomo prikazali le nekaj primerov manj uspešne ali neuspešne segmentacije. Na sliki 3 vidimo primer manj uspešne segmentacije v eksperimentu z grobo označenimi regijami ROI, na sliki 4 pa je prikazan podoben primer le da tokrat za eksperiment s podrobno označenimi regijami zanimanja. V obeh prikazih lahko na levi sliki opazujemo razliko med ročno označenimi in s pomočjo nevronske mreže detektiranimi regijami ROI. Modra barva v tem levem prikazu predstavlja napačno detektirane regije (ang. False Positive), oranžna pa zgrešene regije (ang. False Negative).

Obe naučeni nevronske mreže smo preizkusili tudi na podatkih, ki niso bili del učnih in/ali testnih zbirk, uporabljenih v obeh eksperimentih. S pomočjo vizualnega pregleda dobljenih rezultatov smo zaznali dobro zmožnost generaliziranja obeh naučenih nevronske mreže. Na sliki 5 je prikazan primer segmentiranja slike iz razreda tkiv HP, pri čemer ta slika (slika 5, levo) ni bila del podatkovnih zbirk, uporabljenih v obeh eksperimentih. Na sredinski sliki vidimo dobljen rezultat s pomočjo nevronske mreže, naučene z grobo označenimi regijami zanimanja, na desni sliki pa je prikazan segmentacijski rezultat še za nevronske mreže, naučene s podrobno označenimi regijami zanimanja. Naj ob tem še poudarimo, da slik iz razreda tkiv HP nismo niti uporabili v postopku učenja te mreže.



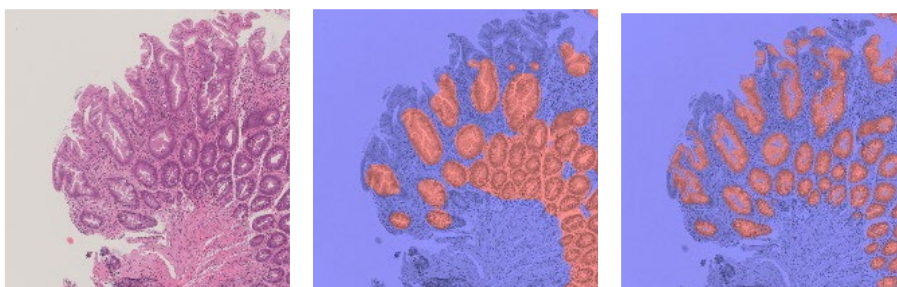
Slika 3: Primer slabo ocenjene segmentacije v eksperimentu z grobo označenimi cilji: slika razlik med oznakami in napovedjo (levo), grobo označene regije zanimanja (sredina) in napoved nevronske mreže (desno). Modra barva v sliki razlik predstavlja napačno detektirane regije (ang. False Positive), oranžna pa zgrešene regije (ang. False Negative).

Vir: lasten



Slika 4: Primer slabo ocenjene segmentacije v eksperimentu s podrobneje označenimi cilji: slika razlik med oznakami in napovedjo (levo), grobo označene regije zanimanja (sredina) in napoved nevronske mreže (desno). Modra barva v sliki razlik predstavlja napačno detektirane regije (ang. False Positive), oranžna pa zgrešene regije (ang. False Negative).

Vir: lasten



Slika 5: Primer segmentacije slike iz razreda tkiv HP, ki ni bil del učne ali testne zbirke obeh eksperimentov: originalna histopatološka slika (levo), napoved nevronske mreže iz prvega eksperimenta (sredina) in napoved nevronske mreže iz drugega eksperimenta (desno).

Vir: lasten

5 Zaključek

V članku smo opisali pripravo podatkov in učenje modela za segmentacijo regij zanimanja s pomočjo dve oblik pripravljenih oznak: grobo in podrobno označene regije zanimanja. Grobo smo označili 123 slik, za kar smo potrebovali približno 3 človek dni. Podrobno pa smo označili samo 10 slik, za kar smo potrebovali približno 1,5 človek dni.

Kljub majhnemu številu označenih slik sta se oba modela uspešno naučila segmentirati ciljne regije zanimanja. Za učenje s podrobno označenimi regijami smo potrebovali zelo malo slik, vendar smo za označevanje porabili bistveno več časa na sliko.

Ovrednoteni rezultati ne kažejo bistvenih razlik, oba modela imata namreč zelo primerljivo natančnost in priklic. Obe metriki sta zadovoljivo visoki. Natančnost je v obeh eksperimentih nad 0,95 in priklic pa nad 0,85.

Če podrobneje pogledamo rezultate modela, učenega z grobo označenimi regijami, opazimo na sliki 4, da ta model pravzaprav v veliko primerih bolj natančno segmentira regije, kot so bile označene. V predikciji je tudi veliko zgrešenih regij, kar pa negativno vpliva na ocenjen priklic metode. Naučena segmentacija je tako bližje podrobnim oznakam, vendar še vedno ne popolnoma natančna.

Kvalitativni rezultati kažejo, da oba modela dokaj dobro generalizirata na preostale slike podatkovne zbirke (glej sliko 5). Model učen z grobo označenimi segmenti se je zmožen naučiti segmentirati veliko bolje, kot pa so bile uporabljene ciljne oznake v fazi učenja. Na koncu lahko povzamemo, da pa model, naučen z bistveno manj, a hkrati podrobno označenimi učnimi primeri, segmentira še bolje od prej omenjenega modela.

Oba naučena modela torej uspeta detektirati regije zanimanja, ki so pomembne za določitev tipa tkiva v histopatoloških posnetkih kolorektalnih polipov. V nadaljnjih raziskavah se bomo osredotočili na vprašanje, ali je mogoče oba modela in značilnice, ki jih luščita oba modela, uporabiti za izboljšanje razvrščanja tkiva.

Viri in literatura

- Barbano, C. A., Perlo, D., Tartaglione, E., Fiandrotti, A., Bertero, L., Cassoni, P., & Grangetto, M. (2021). Unitopatho, a labeled histopathological dataset for colorectal polyps classification and adenoma dysplasia grading (str. 76-80). *2021 IEEE International Conference on Image Processing (ICIP)*.
- Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database (pp. 248-255). In *2009 IEEE conference on computer vision and pattern recognition*.
- K. Simonyan, & A. Zisserman. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Kingma, D. P., & Ba, J. (2015). Adam: A Method for Stochastic Optimization. *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA*.
- Madabhushi, A., & Janowczyk A. (2016). Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases (str. 7-29). *Journal of pathology informatics*.
- Ronneberger, O., Fischer, P., & Brox, T. (2015). 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* (pp. 234-241). Springer International Publishing.
- Song, Z., Yu, C., Zou, S., Wang, W., Huang, Y., Ding, X., ... & Shi, H. (2020). Automatic deep learning-based colorectal adenoma detection system and its similarities with pathologists. *BMJ open*, 10(9), e036423.
- Šavc, M., & Potočnik B. (2023). Primerjava osnovnih konvolucijskih nevronskih mrež za razvrščanje histopatoloških posnetkov kolorektalnih polipov (str. 399-402). *Zbornik dvaintridesete mednarodne Elektrotehniške in računalniške konference ERK 2023*.

UPORABA METOD RAČUNALNIŠKEGA VIDA ZA POMOČ PRI TELESNI VADBI

BLAŽ ČERNI, BORUT BATAGELJ

Univerza v Ljubljani, Fakulteta za računalništvo in informatiko, Ljubljana, Slovenija
blaz.cerni@hotmail.com, borut.batagelj@fri.uni-lj.si

Pred nekaj leti smo bili priča hudi krizi, povezani s koronavirusom, ki je močno vplivala tudi na človeške navade pri telovadbi in izvajanju aktivnosti na prostem ali v fitnes centrih. Zaradi nezmožnosti obiskovanja teh prostorov se je vse več ljudi odločalo za različne oblike vadbe doma. Pri tem se je pojavilo vprašanje pravilne in varne izvedbe vadbe, še posebej pri začetnikih. V tem delu analiziramo, kako nam lahko metode računalniškega vida pomagajo pri prepoznavanju človeške drža in telesnih delov ter kako lahko te informacije uporabimo za usmerjanje in pomoč pri pravilni izvedbi telesnih vaj. Predstavimo tudi našo implementacijo aplikacije, imenovane Sweatpose, ki omogoča zaznavanje in sledenje telesnih delov ter hkrati uporabniku nudi pomoč pri pravilni izvedbi vaj.

DOI
<https://doi.org/10.18690/um.feri.2.2025.6>

ISBN
978-961-286-960-1

Ključne besede:

računalniški vid,
človeška drža,
fitnes,
ključne točke človeškega
telesa,
tensorflow

Prispevek temelji na:

Černi, B. (2024). *Uporaba metod računalniškega vida za pomoč pri telesni vadbi*. magistrsko delo, Univerza v Ljubljani, Fakulteta za računalništvo in informatiko.



Univerzitetna založba
Univerze v Mariboru

DOI
[https://doi.org/
10.18690/um.feri.2.2025.6](https://doi.org/10.18690/um.feri.2.2025.6)

ISBN
978-961-286-960-1

Keywords:

computer vision,
pose estimation,
fitness,
human body key points,
tensorflow

The proceedings is based on: Černi, B.(2024). *Uporaba metod računalniškega vida za pomoč pri telesni vadbi*. master's thesis, University of Ljubljana, Faculty of Computer Science and Informatics.

USE OF COMPUTER VISION METHODS FOR GUIDANCE IN PHYSICAL EXERCISES

BLAŽ ČERNI, BORUT BATAGELJ

University of Ljubljana, Faculty of Computer and Information Science, Ljubljana,
Slovenia
blaz.cerni@hotmail.com, borut.batagelj@fri.uni-lj.si

A couple of years ago, we witnessed a severe crisis related to the coronavirus. This crisis significantly affected people's exercise habits and activities, whether outdoors or in fitness studios. Due to the inability to visit such places for exercise, more and more people opted for various forms of home workouts. This posed a challenge regarding the proper and safe execution of these exercises, especially for beginners. In this work, we analyze how computer vision methods can assist in recognizing human posture and key body points and how this information can guide and support individuals in performing exercises correctly. Finally, we present our implementation of such an application, which, with the help of pose estimation, provides users with assistance in executing their workouts properly. We named the application Sweatpose.



1 Uvod

Metode računalniškega vida, ki lahko na podlagi modelov ocenjujejo ključne točke človeškega telesa in človeško držo, se imenujejo metode za ocenjevanje človeške drže (angl. *pose estimation*). Gre za kompleksno področje v računalniškem vidu, ki omogoča napravam natančno določanje položaja anatomskih točk, kot je na primer lokacija kolena neke osebe na sliki. Rezultati ocenjevanje človeške drže, so napovedane ključne tičke, ki so označene z enoličnim identifikatorjem in največkrat vsebujejo tudi oceno zaupanja med 0 in 1. V zadnjih letih se je razvilo več različnih pristopov za ocenjevanje človeške drže. Nekateri pristopi se osredotočajo samo na določene dele telesa (npr. roke ali stopala), drugi pa na ocenjevanje človeške drže za samo eno osebo ali pa več oseb hkrati. Pri slednjih se pojavijo večji izzivi zaradi naraščajoče kompleksnosti, kar je posledica večjega števila oseb. Med osebami pa lahko prihaja tudi do prekrivanja (Černi, 2024).

Za implementacijo metod, ki ocenjujejo človeško držo večih oseb hkrati, sta znana predvsem dva načina implementacije:

- Pristop od zgoraj navzdol (angl. *top-down approach*). Ta metoda vključuje detektor oseb, ki najprej prepozna osebe in lokacijo njihovega telesa na sliki, potem pa se ocenjevanje ključnih točk telesa izvede za vsako osebo posebej.
- Pristop od spodaj navzgor (angl. *bottom-up approach*). V nasprotju s pristopom od zgoraj navzdol, pa ta metoda prepozna vse dele telesa na sliki naenkrat in potem poveže dele telesa z ustreznimi osebami.

Pristop od zgoraj navzdol je lažji za implementacijo, ker je bolj sistematičen in direktno uporablja tehnike pristopa za ocenjevanje ključnih točk ene osebe. Slaba stran pa je, da je odvisen od uspešnosti detekcije oseb na sliki. Poleg tega pa se čas za izvedbo celotnega algoritma povečuje proporcionalno s številom oseb na sliki kar posledično onemogoča dobro pretočnost v realnem času. Iz tega vidika je pristop od spodaj navzgor v splošnem sprejet kot boljša opcija, ker omogoča visoko natančnost in pretočnost v realnem času, ne glede na število oseb na sliki.

Po uspešni identifikaciji oseb s pomočjo enega izmed zgoraj opisanih pristopov pa model za ocenjevanje človeške drže uporabi pridobljene koordinate, da sestavi predstavitev položajev ključnih točk. Ta predstavitev je lahko v 2D ali pa v 3D.

Prednost predstavitve v 2D je nižja računska zahtevnost, medtem ko nam ocenjevanje človeške drže v 3D prostoru, doda še koordinato Z (poleg X in Y), k napovedani lokaciji ključne točke, kar nam da informacijo o oddaljenosti osebe od kamere.

Sicer pa poznamo tri glavne tipe modelov za ocenjevanje človeške drže (Slika 1):

- Model na osnovi ogrodja (angl. *Skeleton-Based model*), v katerem ključne točke predstavlja množica sklepov in orientacij udov človeškega telesa. Ta model se lahko opiše tudi kot graf, kjer so vozlišča sklepi, poti med vozlišči pa so udi, ki povezujejo te sklepe.
- Model na osnovi obrisa (angl. *Contour-Based model*), v katerem množica ozbriso zajame povezavo delov telesa (značilno za 2D modele). Takšna predstavitev se je precej uporabljala v zgodnjih aktivnih modelih oblike (angl. *Active Shape Models*) ali krajše ASM (T.F. Cootes, 1995).
- Model na osnovi volumna (angl. *Volume-Based model*). Je naprednejši model, v katerem se človeško telo predstavi kot 3D volumen. Moderne predstavitve predstavlja mrežna oblika, ki se pridobi iz 3D skanov. Široko uporabljena modela na osnovi voluma sta SCAPE (Dragomir Anguelov, 2005) in SMPL (Matthew Loper, 2015).

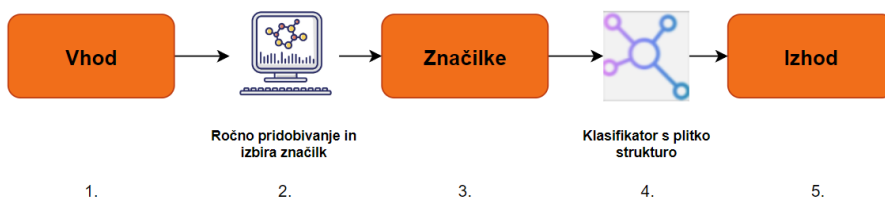


Slika 1: Tipi modelov za ocenjevanje človeške drže od leve proti desni: model na osnovi ogrodja, model na osnovi obrisa in model na osnovi volumna.

Vir: (Yucheng Chen, 2020).

2 Pregled področja

Začetni pristopi za ocenjevanje človeške drže temeljijo na tradicionalnih algoritmičnih strojnega učenja, ki tipično sledijo toku, ki je prikazan na sliki Slika 2. Diagram se začne z vhodnimi podatki (1.), v našem primeru gre za slike. V drugem koraku (2.) sledi ročno pridobivanje in izbiranje značilnk, ki so pomembne za obdelavo podatkov. Po obdelavi vhodni podatki postanejo značilke (3.), kar pomeni, da so pretvorjeni v številčne ali strukturirane oblike. Značilke nato preidejo skozi preprost klasifikator (4.), ki ima plitvo strukturo (na primer SVM ali odločitvena drevesa). Na koncu pa dobimo napoved oziroma izhod (5.), ki je lahko razvrstitev v različne kategorije. Postopek je potraten predvsem zaradi ročnega luščenja značilnk.

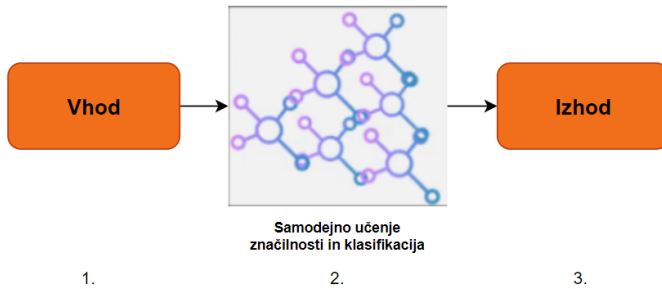


Slika 2: Postopek klasičnega pristopa za ocenjevanje človeške drže.

Vir: lasten.

Klasične pristope so uporabili avtorji v delu, kjer so poskušali problem reševati kot klasični problem s pomočjo naključnih gozdov (G. Rogez, 2008). Drugi (Felzenszwalb, 2005) pa so se ocenjevanja človeške drže lotili s pomočjo koncepta »Slik struktur« (angl. *Pictorial Structures*). Problem so zastavili kot optimizacijski problem, katerega cilj je bil najti najboljšo razporeditev in ujemanje različnih območij slik z videzom določenega dela telesa ter čim boljše geometrično razmerje med deli telesa (npr. bližina glave od preostalega dela telesa).

Iz tega razloga so se pristopi hitro osredotočili na umetne nevronske mreže in globoko učenje. Uteži na posameznih nevronskih povezavah, se tako preračunavajo na podlagi objektov in značilnk vhodne slike in znajo tudi razlikovati med njimi. Posebne in skrite vzorce lahko odkrijejo brez potrebnega predprocesiranja.



Slika 3: Postopek ocenjevanja človeške drže s pomočjo globokega učenja.

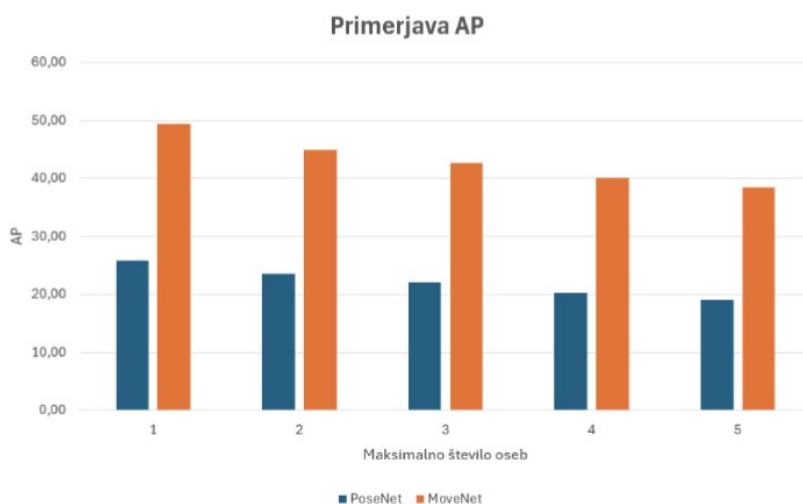
Vir: lasten.

Modeli, ki temeljijo na globokem učenju in nevronske mrežah, tipično prekosijo tradicionalne modele strojnega učenja v nalogah računalniškega vida kot je na primer ocenjevanje človeške drže. Med takšne modele štejemo Mask R-CNN (K. He, 2017), kjer gre v resnici za splošno ogrodje za detekcijo in segmentacijo objektov na sliki, katero pa omogoča tudi ocenjevanje človeške drže. Drugi znani modeli, ki temeljijo na globokem učenju so, AlphaPose (Hao-Shu Fang, 2023) in pa eden izmed najbolj znanih modelov s pristopom od spodaj navzgor OpenPose (Z. Cao, 2017). Postopek ocenjevanja človeške drže s pomočjo globokega učenja prikazuje Slika 3.

3 Izbor modela

Analiza modelov in pregled področja sta razkrila, da obstaja precej veliko različnih načinov in pristopov, s pomočjo katerih se lahko implementira aplikacija, ki s pomočjo ocenjevanja človeške drže, nudi ljudem pomoč in usmeritve pri izvajanju telesnih vadb. Za našo aplikacijo se osredotočimo na modele iz knjižnice Tensorflow.js (LeViet & Chen, 2021). Gre za odprtokodno programsko knjižnico za strojno učenje in umetni inteligenco, kjer za namen ocenjevanja človeške drže ponujajo modele MoveNet, PoseNet in BlazePose. Knjižnica omogoča izvajanje modela na strani odjemalca (na odjemalčevi strojni opremi), v sklopu spletne aplikacije. Vsi izmed naštetih modelov podpirajo izvajanje na grafični procesni enoti (GPU), kjer se v ozadju uporablja tehnologija WebGL ali pa WebGPU, prav tako pa tudi na centralni procesni enoti (CPU), z uporabo WebAssembly. Gre za kodo, ki je »bližje« strojnemu jeziku in zato omogoča precej boljše delovanje kot navadna verzija CPU in je v določenih primerih primerljiva z GPU verzijami (predvsem za primer uporabe pametnih telefonov). Za enega izmed ciljev pri implementaciji aplikacije si zastavimo podporo izvajanja vadbe več osebam hkrati, zato izmed izbora modelov

izločimo BlazePose, ki omogoča samo izvajanje za eno osebo. Med modeloma MoveNet in PoseNet pa se odločimo po izvedeni evalvaciji modelov na COCO podatkovni zbirki (Tsung-Yi Lin, 2014), kjer za naš primer uporabe iz zbirke 5000 testnih slik izberemo samo tiste, na katerih so ljudje. Natančneje, na katerih je ena (1045 slik), dve (1481 slik), tri (1749 slik), štiri (1897 slik) ali pa maksimalno pet oseb (2016) slik. Za ocenjevalno metriko izberemo uradno COCO metriko – povprečno točnost (AP – angl. *average precision*), ki temelji na metriki OKS (angl. *Object KeyPoint Similarity*).



Slika 4: Povprečna točnost (AP) za modela MoveNet in PoseNet, glede na naraščajoče maksimalno število oseb na slikah.

Vir: lasten.

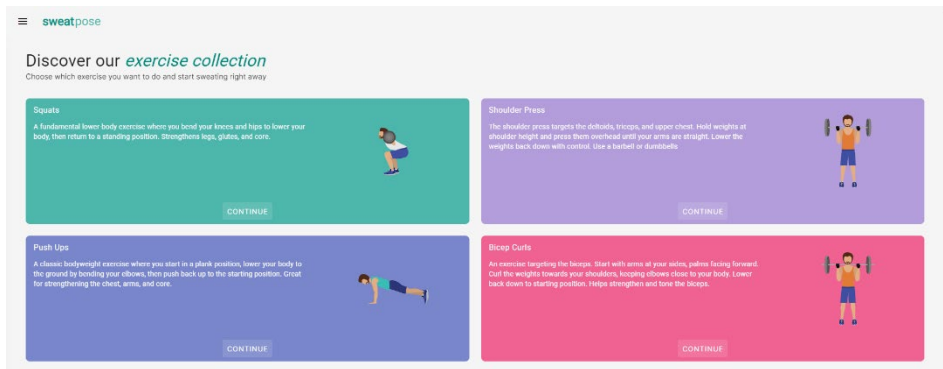
Rezultat evalvacije predstavlja Slika 4, na kateri vidimo, da model MoveNet dosega boljše rezultate zato, ga tudi izberemo za implementacijo aplikacije.

4 Aplikacija Sweatpose

V sklopu aplikacije Sweatpose, ki bo uporabnikom nudila pomoč in usmeritve pri izvajanju telesnih vadb podpremo štiri različne tipe telesnih vadb, ki jih lahko uporabniki izvajajo (Slika 5):

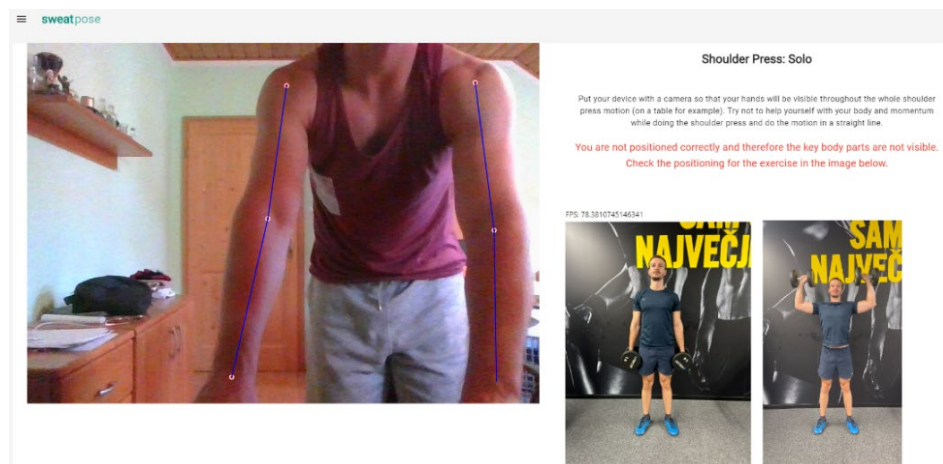
- Počepi – angl. *Squats*.

- Potisk nad glavo – angl. *Shoulder press* ali *Overhead press*.
- Skleca – angl. *Push ups*.
- Vzgib bicepsa – angl. *Bicep curls*.



Slika 5: Zaslonska slika zbirke telesnih vadb, ki jih podpremo v aplikaciji Sweatpose.

Vir: lasten.



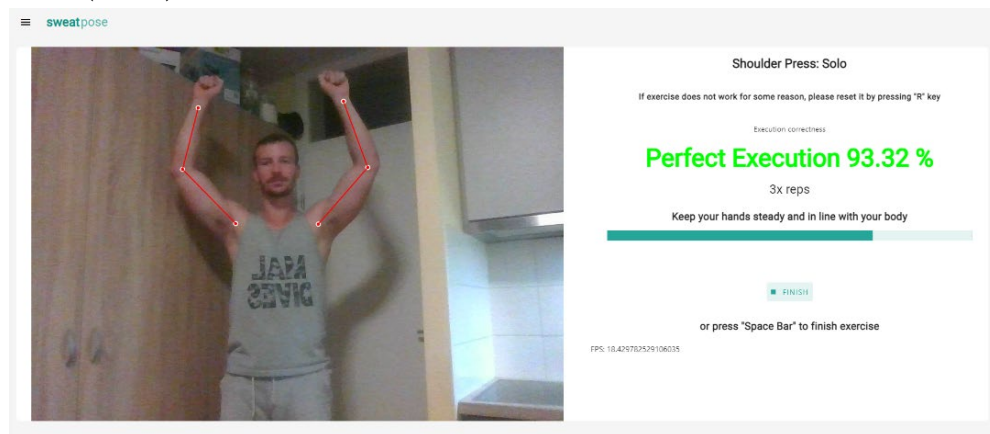
Slika 6: Prikaz za uporabnika pred začetkom vadbe, kjer je na levi strani slika iz kamere, na desni pa navodila: informacija o opravljeni oziroma nepravilni postavitvi, slika pravilne postavitve in posnetek pravilne izvedbe izbrane vadbe.

Vir: lasten.

Vsako vadbo je možno izvajati v tekmovalnem načinu dveh uporabnikov ali pa v načinu posameznika. Prav tako se izbere strojna oprema, na kateri se izvaja model za ocenjevanje človeške drže – grafična procesna enota ali centralna procesna enota

v kombinaciji z WebAssembly. Pred začetkom vadbe se mora uporabnik tudi pravilno postaviti pred kamero. Pomembno je, da so vidne vse ključne točke telesa, ki so za to vadbo glavne. Aplikacija uporabnika o tem tudi obvesti vizualno in tekstovno (Slika 6).

Ko uporabnik v pravilnem položaju stoji nekaj sekund, se vadba začne. Uporabniku se nato začnejo prikazovati informacije o natančnosti izvedbe vadbe v odstotkih in obarvanosti obvestila, število izvedenih ponovitev vadbe, grafični prikaz izvedbe posamezne ponovitve in usmeritve oziroma opozorila za bolj pravilno izvedbo vadbe (Slika 7).



Slika 7: Izvedba vadbe s prikazom slike iz kamere in prepoznanih delih telesa na levi strani in povratni informaciji o natančnosti izvedbe vadbe na desni strani.

Vir: lasten

Ko ima uporabnik dovolj vadbe, jo lahko konča s klikom na gumb »Finish« oziroma s tipko preslednice na tipkovnici. Po koncu vadbe se uporabniku prikaže pojavno okno, kamor lahko vpiše podatek o uporabljenih dodatnih utežeh oziroma svoji teži (na primer pri sklecah), za pomoč pri izračunu kalorij. Po potrditvi dodatnih informacij se uporabniku prikaže stran s povzetkom vadbe.

4.1 Implementacija izvajanja telesnih vadb

Implementacijo ločimo na tri dele, ki se izvajajo na čelnem delu (angl. *frontend*), kar pomeni, da jih poganja strojna oprema uporabnika (odjemalca):

- Inicializacija kamere in modela pred začetkom vadbe.
- Preverjanje ustreznosti postavitve uporabnika pred začetkom vadbe z uporabo modela MoveNet.
- Uporaba modela MoveNet med izvajanjem vadbe za pomoč in usmerjanje uporabnika k pravilni izvedbi vadbe in izračunavanje povratnih informacij, kot je točnost izvedbe v odstotkih.

Ob začetku posamezne vadbe se najprej pridobi zajem slike iz kamere. Kamera prenaša sliko v realnem času in jo hrani v HTML elementu *video*. Za dejanski prikaz uporabniku, pa uporabimo HTML element *canvas*, na katerega lahko potem tudi rišemo. Po inicializaciji kamere, se zgodi še inicializacija modela MoveNet. Glede na izbran tip strojne opreme, se pridobijo ustrezne definicije modelov iz oddaljenega Tensorflow repozitorija. V primeru izbora grafične procesne enote za način izvajanja, se pridobijo definicije modelov za izvajanje na grafični procesni enoti s podporo WebGPU (WebGPU, 2024). Gre za JavaScript API, ki ga podpirajo novejšje verzije brskalnikov in omogoča učinkovito izrabo grafične procesne enote. V primeru izbora centralne procesne enote, pa se pridobijo definicije modelov za izvajanje modelov s podporo WebAssembly, kjer pa gre za nizkonivojsko kodo, kar omogoča učinkovito izrabo strojne opreme. Inicializiran model vsebuje tudi sledilec (angl. *tracker*), ki omogoča sledenju več osebam, preko več zaporednih sličic. Vsaki zaznani osebi se na začetku priredi unikaten identifikator, s pomočjo katerega se skozi zaporedne sličice omogočata sledljivost in razumevanje, kdo je kdo.

V fazi preverjanja ustreznosti postavitve uporabnika pred kamero za izvajanje posamezne vadbe, model MoveNet že pridobiva ključne točke človeškega telesa. Tiste, ki so v uporabi se uporabniku tudi izrišejo in prikažejo. Faza preverjanja ustreznosti postavitve je pomembna iz dveh razlogov. Prvi je ta, da se takrat izvedejo določeni začetni izračuni oziroma začetne pozicije ključnih točk človeškega telesa, ki so potem v uporabi pri samem izvajanju vadbe. Drugi razlog pa je precej bolj praktične narave. Gre za to, da ko je uporabnik postavljen v pravilni, začetni drži za izvajanje določene vadbe, se dejansko začne izvedba vadbe, in ocenjevanje uporabnikovega izvajanja vadbe, na primer sklec. Aplikacija Sweatpose pravilnost postavitve pred začetkom vadbe, preverja na nivoju posamezne vadbe. Za vsako, posamezno vadbo sistem pričakuje določen položaj uporabnika oziroma bolje rečeno, določene ključne točke, ki morajo biti za začetek vadbe vidne in z dovolj

visoko oceno ocenjene (gre za oceno, ki jo ob zaznavanju ključnih točk vrača model MoveNet).

Primer takšnih pravil za izvajanje vadbe Potisk nad glavo:

- Vidnost in dovolj visoka ocena točnosti ključnih točk – zapestja na levi in desni roki, komolca na levi in desni roki in ramen.
- Pomembna je usmerjenost uporabnika neposredno v smer kamere.

Pravilna postavitve pred vadbo je izjemnega pomena za točnost ocenjevanja drže in ključnih točk s pomočjo modela MoveNet. Ker gre za model, ki ocenjuje v 2D, mora biti telo postavljeno tako, da omogoči pravilen izračun kotov med ključnimi točkami, ki se potem uporabljajo za ocenjevanje točnosti izvedbe vadbe.

Po pravilni postavitvi uporabnika pred kamero, se začne dejanska izvedba vadbe iz uporabnikovega pogleda. Model MoveNet na tej točki že izvaja ocenjevanje ključnih točk in drže človeškega telesa. Informacije o teh točkah uporabimo v sistemu, ki uporabniku nudi povratne informacije o tem kako izvaja posamezno vadbo. Povratne informacije so sestavljene iz procentualne točnosti izvedbe vadbe, števila ponovitev posamezne vadbe in predlogov ter opozoril, čemu se pri izvedbi vadbe izogibati in kako vadbo izvajati za boljši rezultat. Za vsako izmed štirih vadb implementiramo ločena pravila, ki se preverjajo za posameznega uporabnika. V ta namen najprej definiramo ključne točke, ki so pomembne za izvedbo vadbe posameznega tipa. Definirane ključne točke nato uporabimo za izračun kotov med njimi. Poleg kotov, ki jih izračunamo, za vsako vadbo definiramo tudi druge ključne spremenljivke, ki hranijo stanje izvedbe posamezne vadbe in parametre kot so na primer začetni in končni položaj vadbe, seznam vseh izračunanih procentualnih točnosti za vsako sličico zajetega posnetka ter število ponovitev izvedene vadbe. Za vsako izmed pravil izračunamo oceno točnosti v intervalu od 0 do 100. Ocene točnosti za pravila, se izračunajo na podlagi odstopanj kotov od idealnih kotov oziroma odstopanj koordinat od pričakovanih koordinat. Na podlagi ocen točnosti posameznih pravil, izračunamo končno oceno točnosti trenutne izvedbe vadbe v trenutni sličici, ki je enaka povprečju ocen točnosti posameznih pravil za nek tip vadbe.

Vadba Potisk nad glavo, je vadba, ki se jo primarno izvaja s pomočjo dodatnih uteži, ki se jih dviguje nad glavo. Aplikacija Sweatpose predpostavlja stoječo verzijo izvedbe vadbe, takšno kot prikazuje Slika 8.



Slika 8: Potisk nad glavo (angl. *shoulder press* ali *overhead press*) v stoje in z rdečo označen kot ter ključne točke, ki jih uporablja aplikacija za preverjanje pravilnosti izvedbe.

Vir: lasten.

Za to vadbo definiramo spodnje ključne točke človeškega telesa:

- Levo in desno zapestje.
- Levi in desni komolec.
- Leva in desna rama.

Definiramo torej šest ključnih točk človeškega telesa, pri ocenjevanju pa smo pozorni na to, da uporabnik hkratno premika obe roki. Zanimajo nas premiki vseh šest ključnih točk naenkrat. Definiramo tudi začetni in končni pričakovan kot izvedbe vadbe. Gre torej za kot, ki ga oklepajo zapestje, komolec in rama z vrhom v ramenu. Potem definiramo pravila na podlagi katerih se izračunavamo procentualno točnost izvedbe in nudimo nasvete in usmeritve uporabniku:

- Na podlagi x koordinat zapestij in komolcev preverjamo oziroma sledimo, da uporabnik z rokami ne zamahuje preveč levo ali desno in da ima roke

približno poravnane s telesom. Na takšen način lahko naredi potisk bolj mirno in naravnost navzgor. V primeru, da temu ni tako, uporabniku sporočimo, da ne drži rok dovolj mirno in v liniji s telesom.

- Na podlagi kotov obeh rok in definiranih začetnih in končnih kotov, preverjamo, da uporabnik ob potisku navzgor ne zravna rok preveč oziroma da rok ob spustu ne spusti prenizko. V primeru, da potiska preveč navzgor ali spušča prenizko, se uporabniku sporoči ustrezna povratna informacija.
- Če uporabnik izvaja vadbo pravilno in stoji pri miru, se y-koordinata ramen med izvedbo vaje ne bi smela bistveno spreminjati. Zato preverjamo spremembe y-koordinate med zaporednimi sličicami. Če je sprememba večja od določene meje, sklepamo, da si uporabnik pri potisku pomaga tako, da se dvigne na prste in s tem pridobi dodatno moč za dvig. V tem primeru ga opozorimo.
- Med izvedbo vaje primerjamo kota obeh rok. Pri pravilni izvedbi se kota ne bi smela bistveno razlikovati, saj uporabnik uteži nad glavo potiska hkrati in usklajeno z obema rokama. Če odstopanje presega določeno mejo, uporabniku podamo povratno informacijo, da mora roke dvigovati bolj usklajeno.

Na podoben način določimo ključne točke in pravila tudi za preostale tipe vadb.

5 Rezultati in ugotovitve

Ena najpomembnejših lastnosti aplikacije za pomoč pri telesni vadbi je njena zadostna pretočnost oziroma delovanje v realnem času med izvajanjem vadbe. Zato izmerimo število sličic na sekundo (FPS) na različnih napravah in strojni opremi med vadbo. Rezultate prikazuje Tabela 1.

Tabela 1: Povprečno število sličic na sekundo (FPS), na različni strojni opremi in različnih napravah.

Naprava	Strojna oprema	FPS
PC	CPU	5
PC	CPU z WebAssembly	20
PC	Intel HD Graphics GPU	15
PC	NVIDIA GeForce RTX 2060 GPU	80
iOS pametni telefon	CPU z WebAssembly	20
Android pametni telefon	CPU z WebAssembly	15
Android pametni telefon	GPU	8

Vsi primeri, kjer se uporablja grafična procesna enota, temeljijo na WebGPU tehnologiji. Izvajanje s pomočjo WebGPU tehnologije je sicer možno na kakršnih koli grafičnih procesnih enotah. Ni nujno, da gre za ločeno grafično procesno enoto. Tudi integrirane grafične procesne enote, kakršne najdemo v prenosnih računalnikih so podprte. Prav tako tudi tiste, ki so integrirane v modernih pametnih telefonih. WebGPU je v resnici JavaScript API, ki omogoča brskalnikom, da na učinkovit način uporabljajo grafično procesno enoto odjemalca. WebGPU naj bi bil naraven naslednik precej bolj znane tehnologije WebGL saj zagotavlja boljšo združljivost z novjšimi grafičnimi karticami. Ker pa gre tukaj za novejšo tehnologijo, je podpora za WebGPU omejena tudi glede na (novejše) verzije brskalnikov, ki jih uporabljamo. Na spletni strani (Can I use WebGPU?, 2024) lahko preverimo natančno kakšna je trenutno podpora za WebGPU, v različnih brskalnikih. Chrome, Edge in tudi Opera podpirajo WebGPU z verzijami od začetka leta 2023 naprej, medtem ko je podpora omejena na brskalniku Firefox in Safari. Tudi uporaba WebAssembly je omejena z določenimi verzijami brskalnikov. Je pa ta podpora že precej bolj široka kot pa tista za WebGPU, kar lahko preverimo na spletni strani (Can I use WebAssembly?, 2024). Nižje vrednosti dosega le v primeru uporabe centralne procesne enote brez WebAssembly. Tukaj gre za primer, ko bi inicializacija modela padla tako za grafično procesno enoto kot tudi za centralno s pomočjo WebAssembly. Do tega bi lahko prišlo v primeru, da bi WebGPU neuspešno poskušal dostopati do grafične procesne enote uporabnika oziroma v primeru, ko brskalnik ne bi podpiral niti WebGPU niti WebAssembly. V takšni situaciji se model potem izvaja povsem brez dodatnih izboljšav na CPU. Gre za redek primer, do katerega v splošnem niti ne pride. Sicer pa pričakovano najvišje vrednosti FPS, aplikacija dosega na grafični procesni enoti višjega standarda, kot je omenjena NVIDIA GeForce RTX 2060, kjer je vrednost FPS v povprečju okrog 80 slik na sekundo.

Kljub dobremu rezultatu delovanja aplikacije Sweatpose, pa ima le-ta še vedno nekaj pomanjkljivosti oziroma prostora za napredek in izboljšave. Glavna pomanjkljivost je v načinu kako model MoveNet izvaja ocenjevanje ključnih točk človeškega telesa. Ker smo želeli podpreti ocenjevanje človeške drže za več oseb hkrati, smo izbrali MoveNet model, ki pa v tem primeru izračunava in rezultate podaja v 2D koordinatah, kar so v resnici x in y koordinate pikslov na vsaki sličici zajema iz kamere. Tukaj se potem izgubi globina oziroma oddaljenost oseb od kamere. To se je ob definiranju pravil za izvajanje vadb, kjer smo upoštevali na primer razlike med koordinatami in podobno, izkazalo za pomembno pomanjkljivost. Pravilna postavitev pred kamero je zaradi pomanjkanja 3D koordinat v prostoru ključnega

pomena, prav tako lahko učinkovitost ocenjevanja točnosti izvedbe vadb malce varira glede na oddaljenost osebe, ki je postavljena pred kamero. Izboljšave bi se lahko lotili tako, da bi vzeli model, ki podpira 3D koordinate za posameznika potem pa bi z dodatnim modelom za detekcijo objektov in oseb na primer YOLO (Joseph Redmon, 2016) detektirali posamezne osebe in na vsaki posamezni posebej pognali model za 3D ocenjevanje ključnih točk človeškega telesa. To pa bi gotovo potegnilo za seboj drug problem, ker bi bilo težje takšno zadevo izpeljati v realnem času.

6 Zaključek

V delu predstavimo kaj je ocenjevanje človeške drže s pomočjo metod računalniškega vida in v kakšne namene ga uporabljamo. Izvedemo analizo različnih pristopov (klasični, globoko učenje, od spodaj navzgor, od zgoraj navzdol) in modelov za ocenjevanje človeške drže (model na osnovi ogrodja, model na osnovi obrisa in model na osnovi voluma). Za iskanje ključnih točk človeškega telesa, je najbolj uporaben model na osnovi ogrodja, ker je ravno ta sestavljen iz množice ključnih točk (po navadi sklepov) človeškega telesa, ki se lahko potem uporabljajo v 2D ali 3D predstavitevah.

Na podlagi analize in pridobljenega znanja izberemo primeren model za implementacijo aplikacije, ki bo uporabnikom nudila pomoč in usmeritve pri izvedbi fizičnih vadb. Izberemo model MoveNet iz zbirke Tensorflow.js in z uporabo tega razvijemo aplikacijo Sweatpose, ki omogoča izvedbo štirih različnih tipov vadb (počepi, skleca, potisk nad glavo in vzgib bicepsa), na različni strojni opreми (GPU ali CPU) in različnih napravah. Med izvedbo vadbe uporabnik v realnem času dobiva povratne informacije in nasvete o izvedbi vadbe. Izvajanje v realnem času pa tudi izmerimo s pomočjo vrednosti števila sličic na sekundo (FPS), hkrati pa izpostavimo tudi obstoječe pomanjkljivosti in možne izboljšave.

Viri in literatura

- Černi, B. (2024). Uporaba metod računalniškega vida za pomoč pri telesni vadbi, magistrsko delo, Univerza v Ljubljani, Fakulteta za računalništvo in informatiko.
- Can I use WebAssembly?* (2024). Pridobljeno iz <https://caniuse.com/wasm>
- Can I use WebGPU?* (2024). Pridobljeno iz <https://caniuse.com/webgpu>
- Dragomir Anguelov, P. S. (2005). SCAPE: Shape Completion and Animation of People. *ACM Trans*, 408-416. doi:<https://dl.acm.org/doi/10.1145/1073204.1073207>
- Felzenszwalb, P. H. (2005). Pictorial Structures for Object Recognition. *International Journal of Computer Vision*, 55-79. doi:<https://doi.org/10.1023/B:VISI.0000042934.15159.49>

- G. Rogez, J. R. (2008). Randomized trees for human pose detection. *IEEE Conference on Computer Vision and Pattern Recognition*, 1-8. doi:10.1109/CVPR.2008.4587617
- Hao-Shu Fang, J. L. (2023). AlphaPose: Whole-Body Regional Multi-Person Pose Estimation and Tracking in Real-Time. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7157-7173. doi:10.1109/TPAMI.2022.3222784
- Joseph Redmon, S. D. (2016). You Only Look Once: Unified, Real-Time Object Detection. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. doi:10.1109/CVPR.2016.91
- K. He, G. G. (2017). Mask r-cnn. *Proceedings of the IEEE international conference on computer vision*, 2961-2969.
- LeViet, K., & Chen, Y.-h. (2021). Pose estimation and classification on edge devices with MoveNet and TensorFlow Lite. *TensorFlow Blog*. Pridobljeno iz tensorflow.org: https://www.tensorflow.org/lite/tutorials/pose_classification
- Matthew Loper, N. M.-M. (2015). SMPL: a skinned multi-person linear model. *ACM Trans*, 16. doi:<https://doi.org/10.1145/2816795.281801>
- T.F. Cootes, C. T. (1995). Active Shape Models-Their Training and Application. *Computer Vision and Image Understanding*, 61(1), 38-59. doi:<https://doi.org/10.1006/cviu.1995.1004>
- Tsung-Yi Lin, M. M. (2014). Microsoft COCO: Common Objects in Context. (str. 740-755). Springer International Publishing.
- WebGPU. (2024). Pridobljeno iz <https://www.w3.org/TR/webgpu/>
- Yucheng Chen, Y. T. (2020). Monocular human pose estimation: A survey of deep learning-based methods. *Computer Vision and Image Understanding*. doi:<https://doi.org/10.1016/j.cviu.2019.102897>
- Z. Cao, T. S.-E. (2017). Realtime Multi-person 2D Pose Estimation Using Part Affinity Fields. *IEEE Conference on Computer Vision and Pattern Recognition*, 1302-1310. doi:10.1109/CVPR.2017.143

UNEXPLODED ORDNANCE DETECTION IN HYPERSPECTRAL IMAGES BY USING DEEP NEURAL NETWORKS

MILAN BAJIĆ,¹ BOŽIDAR POTOČNIK²

¹ Zagreb University of Applied Sciences, Department of IT and Computer Sciences,
Zagreb, Croatia
mbajic@tvz.hr

² University of Maribor, Faculty of Electrical Engineering and Computer Science,
Institute of Computer Science, Maribor, Slovenia
bozidar.potocnik@um.si

Unexploded Ordnance (UXO) is a major threat affecting the lives of people in more than 60 countries. This work tests deep neural networks to automatically detect UXO in Hyperspectral Images (HSI). Initially, we constructed our own dataset of 134 HSI cubes divided into three folds: two for training and one for validation. U-Net was selected through preliminary experiments as the most promising detection method among those compared. Customised loss functions were designed for the U-Net, resulting in 3 different models. These models were trained and validated in a supervised manner on our data. The results obtained are very promising with a UXO detection rate of around 70% and an F1 score above 0.8.

DOI
[https://doi.org/
10.18690/um.feri.2.2025.7](https://doi.org/10.18690/um.feri.2.2025.7)

ISBN
978-961-286-960-1

Keywords:

nexploded ordnance,
hyperspectral images,
image segmentation,
convolutional neural
networks,
dataset construction



University of Maribor Press

DOI
[https://doi.org/
10.18690/um.feri.2.2025.7](https://doi.org/10.18690/um.feri.2.2025.7)

ISBN
978-961-286-960-1

DETEKTIRANJE NEEKSPLODIRANIH UBOJNIH SREDSTEV NA HIPERSPEKTRALNIH SLIKAH Z UPORABO GLOBOKIH NEVRONSKIH MREŽ

MILAN BAJIĆ,¹ BOŽIDAR POTOČNIK²

¹ Tehničko Veleučilište u Zagrebu, Katedra za IT in računalništvo, Zagreb, Hrvaška
mbajic@tvz.hr

² Univerza v Mariboru, Fakulteta za elektrotehniko, računalništvo in informatiko, Inštitut
za računalništvo, Maribor, Slovenija
bozidar.potocnik@um.si

Ključne besede:
neeksplodirana ubojna
sredstva,
hiperspektralne slike,
segmentacija slik,
konvolucijske nevronske
mreže,
konstrukcija podatkovne
zbirke

Neeksplodirana ubojna sredstva (UXO) so velika grožnja, ki ogroža življenja ljudi v več kot 60 državah. To delo preizkuša globoke nevronske mreže za samodejno zaznavanje UXO v hiperspektralnih slikah (HSI). Na začetku smo izdelali lasten nabor podatkov iz 134 kock HSI, razdeljenih v tri dele: dva za učenje in en za validacijo. U-Net je bil s predhodnimi poskusi izbran kot najbolj obetavna detekcijska metoda med primerjanimi. Prilagojene funkcije izgube so bile zasnovane za U-Net, s čimer smo dobili 3 različne modele. Ti modeli so bili naučeni in preizkušeni na nadzorovan način na naših podatkih. Dobljeni rezultati so zelo obetavni z uspešnostjo detektiranja UXO okoli 70 % in oceno F1 nad 0,8.



Univerzitetna založba
Univerze v Mariboru

1 Introduction

Hyperspectral Imaging (HSI) is a camera-based technique that captures a dense image series, characterized by its wavelength range and number of channels. The Specim IQ camera utilized in this study possesses a spatial resolution of 512x512 pixels and 204 spectral channels. This camera functions on the principle of an integrated hyperspectral push-broom line scanner, with internal processing occurring during the acquisition process. The results of this process are calibrated reflectance values within the range of 400nm to 1000nm, with a mean spectral resolution of 7nm within this range. However, higher FWHM (i.e., full width at half maximum) values at higher wavelengths generally result in data that is not useful (from 900nm to 1000nm). Hyperspectral images are defined as a set of pixels containing spatial dimensions (m rows and n columns) and spectral information (K wavelength channels). This is commonly referred to as a three-dimensional hyperspectral cube (hypercube), data cube, spectral cube, spectral volume, or data volume. Hyperspectral imaging provides valuable insights into the physical and chemical properties of analysed materials, with the gathered information including geometric and physical characteristics such as size, orientation, shape, texture, and colour. A raw hyperspectral image consists of multiple interconnected sub-images, with each sub-image representing an object's spatial distribution and intensity at a specific wavelength. It is possible to extract an individual spatial image from the hypercube at any point within the system's spectral sensitivity range. Consequently, a hyperspectral image represented as $I(x, y, l)$ can be interpreted either as a standalone spatial image $I(x, y)$ at a given wavelength (l) or as a spectral profile $I(l)$ at each pixel (x, y) . Each pixel within the hyperspectral image contains a unique spectral signature corresponding to its specific location, effectively acting as a fingerprint for identifying its composition.

The proposed research is concerned with the detection of explosive objects. It is estimated that over sixty countries currently possess remnants of Unexploded Ordnance (UXO), (Bajić & Potočnik, 2024). For instance, UXOs from the First World War have been unearthed during archaeological excavations and construction (National University of Public Service, Hungary & Ember, 2021), (Roberts & Williams, 1995). Furthermore, over 100 countries have been affected by explosions at ammunition storage sites, resulting in casualties, environmental damage, mass displacement, and economic disruption (iMMAP-IHF, Humanitarian Access

Response - Monthly Security Incidents Situation Report, November 2018, 2019). It is estimated that there have been tens of thousands of accidental explosions at ammunition depots (Ammunition Storage Area Explosions – EOD Clearance, 2021). In contrast to buried UXOs, surface UXOs can be effectively detected with hyperspectral imaging (Bajić et al., 2013), (Bajić & Bajić, 2021). International efforts to address UXO contamination include initiatives such as the Mine Ban Treaty, which aims to eliminate landmines and promote demining activities worldwide. Organizations such as the United Nations Mine Action Service (UNMAS) collaborate with affected countries to clear contaminated areas and educate on the dangers of UXOs. Furthermore, international funding and partnerships have been established to support technological advancements and training for more effective UXO detection and removal. The presence of explosive remnants that have not been removed thus far continues to pose a significant threat.

Deep neural networks have been employed for many years with great success in the segmentation and classification of multidimensional data (e.g. images, volumes) or for object detection. One such early and extensively deployed neural network is the U-Net network (Ronneberger et al., 2015). U-Net is a well-established and recognized Convolutional Neural Network (CNN) architecture with a left-side contracting phase (i.e., encoder), a bottleneck phase, and an expansive right path (i.e., decoder). This configuration, in conjunction with skip connections from the encoder to the corresponding decoder layer, facilitates the preservation of both spatial and spectral characteristics. U-Net was originally developed for the classification and segmentation of biomedical images. Biomedical and hyperspectral imaging are similar in that both rely on multi-channel imaging, which is why it is so difficult to collect large datasets in either area. U-Net was designed for data with typically 3 channels, whereas hyperspectral images typically have more than 100 channels. The integration of U-Net architectures with HSI has demonstrated efficacy in detecting UXOs. According to (Tuohy et al., 2023), UAV-based HSI can effectively detect surface-level Explosive Remnants of War (ERW), suggesting that deep learning models like U-Net could enhance detection accuracy. This study combined U-Net models with HSI data to detect surficial explosive ordnance using UAV-mounted HSI systems. Authors employed a U-Net to detect artillery and rocket craters in Ukraine, achieving an 89% accuracy rate compared to human marking, thereby demonstrating the model's ability to detect UXO. When applied to

hyperspectral data, U-Net-based segmentation offers a more efficient and safer alternative to conventional methods for UXO and ERW detection.

In this study, a computational method is proposed for the identification of UXOs through the utilization of a U-Net neural network in conjunction with HSI data. The subsequent sections will provide a detailed exposition of the acquisition of the dataset, its annotation, and the division of the data into training and testing sets. Section 3 will provide a concise overview of the computational methods employed and the salient characteristics of the convolutional networks utilized in this study. The ensuing section, i.e. Section 4, will present the experiments, implementation details, metrics used, details about model training and hyperparameters' fine-tuning. The results section will present the quantitative and qualitative results obtained. Finally, the work is concluded with a discussion of potential future directions for research.

2 Dataset

The dataset was created using an experimental Graeco Latin square design (Guthrie, 2020). This was done to minimise the specific spatial sensor position, the influence of light change, and the influence of the surface environment. The surface is characterised by gravel, grass, grassland, grass in sunlight, and bush leaves in shadow.

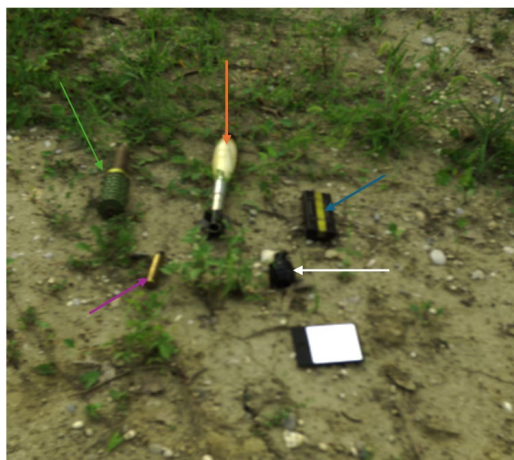


Figure 1: Four UXO, plastic pressure plate, and calibration card

Source: Own.

Four different types of UXO were identified in this study: a mortar mine (marked with an orange arrow in Figure 1), an anti-personnel mine (green arrow), a hand grenade (white arrow), a large bullet (purple arrow) and a plastic pressure plate used to activate explosive ordnance (blue arrow). It is important to note that a white reference card is incorporated into all images, as it is utilised for post-recording calibration in the camera to calculate reflectance values from digital numbers. The approximate image acquisition time ranges between 60 and 120 seconds, depending on exposure and integration settings. The dataset under consideration consists of 134 hyperspectral cubes, with the presence of one to five of the aforementioned objects in every HSI image.



Figure 2: Acquisition of dataset

Source: Own.

The objects were supplied by HCR-CTRO, an educational and research company specialising in land mine clearance education and land mine clearance technology certification. Each of the five objects is present in at least 100 HSI images. The recorded images (see Figure 2) were subjected to a visual control process following the calibration process. In instances where the data proved inadequate, the image capture was repeated with the same position and using the same exposure and integration values. Each cube is stored in the ENVI HDR + data file format. The

data type is 4, interleave BIL, while channels 70, 53 and 19 were utilised for the purposes of visualization and annotation of RGB images.

The images in the database were annotated manually, using visualised RGB images. Each object was denoted by polygons. It was ensured that the same object (i.e., the target) was assigned the same class label in all HSI images. We used the Supervisely platform for annotation. Initial annotations were automatically determined using the 'segment anything' option, which we visually inspected and manually corrected. The resulting mask was used for all wavelengths in the HSI cube.

The dataset under consideration thus contains 134 HSI images. The data was randomly divided into three parts (i.e., subgroups, folds) to enable 3-fold validation. The first two folds comprise 45 images each, while the third fold consists of 44 images. The division of the dataset into subgroups is illustrated in Table 1 for each fold.

Table 1: Dataset division aimed for 3-fold cross validation. The numerical values represent the image ID within the database.

	Image ID
Fold 1	180, 181, 182, 184, 186, 187, 189, 196, 197, 201, 202, 203, 204, 206, 208, 215, 221, 225, 241, 244, 249, 250, 252, 256, 258, 259, 266, 271, 273, 275, 276, 281, 284, 285, 286, 288, 295, 297, 298, 301, 304, 305, 307, 309, 310
Fold 2	178, 179, 183, 190, 195, 198, 211, 212, 214, 219, 220, 222, 223, 226, 227, 232, 237, 238, 239, 240, 243, 246, 251, 254, 255, 260, 265, 268, 269, 270, 272, 274, 280, 282, 283, 287, 291, 293, 296, 299, 300, 302, 311, 312, 313
Fold 3	185, 188, 191, 193, 194, 199, 200, 205, 207, 210, 213, 216, 217, 218, 224, 228, 229, 230, 231, 233, 234, 235, 236, 242, 245, 247, 248, 253, 257, 261, 262, 263, 264, 267, 277, 278, 279, 289, 290, 292, 294, 303, 306, 308

It should be noted that the file size of each HSI image is 209 MB, whereas the mask (i.e., annotations) is 1.36 KB.

3 Computational methods

In the preliminary phase of this research, experimentation was conducted with various CNN architectures incorporating distinct attention mechanisms, utilising a binary cross entropy loss function and an Adam optimiser. The following architectures were assessed: U-Net and U-Net with attention, DeepLabV3+ and

FCN. The aim was to identify the most promising architecture for further experimentation.

U-Net is an exemplary, well-established convolutional neural network that is founded on an encoder-decoder architecture (Ronneberger et al., 2015). Attention U-Net is a modification of U-Net that uses attention gates by upsampling operation at each decoder layer. In skip connections, the gates are used to highlight the important regions and, thus, allowing the network to focus on the more relevant features (Yan et al., 2018). DeepLabV3+ employs Atrous Spatial Pyramid Pooling (ASPP), a technique that captures features at multiple scales. This approach facilitates the model's comprehension of the context. Atrous convolution, a pioneering innovation, introduces gaps between the values in a convolutional kernel, thereby expanding the filter's capacity to encompass larger areas of the input image without increasing the number of parameters (Chen et al., 2018). The replacement of fully connected layers with convolutional layers in a Fully Convolutional Network (FCN) (Long et al., 2015) results in a network capable of accommodating inputs of any size. FCN incorporates skip connections for feature map combination; however, these are generally simpler and less structured in comparison to U-Net.

Table 2 summarizes the trainable parameters and number of layers for the four tested models.

Table 2: Four tested CNNs: Number of layers and trainable parameters

	Model	Trainable parameters	Number of layers
1.	FCN	33.311.928	156
2.	U-Net	34.590.913	81
3.	Attention U-Net	31.977.317	97
4.	DeepLabV3+	10.689.537	434

All tested models were downloaded from the Keras and Pytorch GitHub, respectively. The architectures of the models were not modified, only the input and output layers were adapted to our problem (the expected input size was 512x512x102 and the size of the predicted binary mask was 512x512x1). The Binary Cross Entropy (BCE) Loss function has been utilized. All experiments were performed on an optimised Google Colab pay-per-use environment, in combination with Google Drive for large data storage. The GPU used was an A8 with 40 GB of

RAM, with 81 GB of system RAM. The code was written in Python, and the models were saved in the h5 format.

In preliminary tests, the models were evaluated on only 5 HSI test images, the rest were used for training. Table 3 summarises the metrics obtained, calculated at the pixel level. It can be seen that the most balanced results were obtained using the classic U-Net, and this model was, therefore, selected for all further experiments. In the sequel of this research, we thus experimented with the original U-Net architecture, focusing on the design of a custom loss function.

Table 3: Results of a preliminary model selection experiment. The best model is in bold.

Model	Dice	IoU	Precision	Recall
Attention U-Net	0,639	0,523	0,710	0,640
DeepLabv3+	0,663	0,519	0,851	0,568
U-Net	0,764	0,648	0,917	0,689
FCN	0,992	0,008	0,875	0,632

4 Results

This section describes the experimental design and implementation details, followed by a presentation of the quantitative and qualitative results obtained.

4.1 Experiments and implementation details

We dealt with UXO detection from HSI images in this research. All UXOs were treated as a common class (i.e., we did not distinguish between different types of UXOs). We therefore dealt with the so-called binary segmentation problem: pixel belongs to UXOs or pixel is part of the background. The computational method used was the U-Net neural network (see previous section). Three different modifications of the U-Net are proposed in this work (models M1 to M3). All of these models are based on the same 20-layer architecture, they just use different loss functions.

In this research, we therefore focused on investigating the effect of the loss function on segmentation effectiveness. A custom loss function was designed by combining the following loss functions in the form of a linear combination: Binary Cross Entropy Loss, Dice Loss, Focal Loss, Tversky Loss and Edge Loss. Edge loss is also

a custom loss function that we use to estimate areas near the edges of objects where there should be differences in the spectral signature. The weights of the loss functions have been chosen so that their sum equals 1. All three models had the same weights for BCE loss (set to 0.25), Dice loss (0.4), and Focal loss (0.1). Model M1 had the weights for the Tversky loss set to 0.1 (with $\alpha=0.2$) and the Edge loss set to 0.15, model M2 had these weights set to 0.15 (Tversky loss) and 0.1 (Edge loss), while model M3 used the weight of 0.1 (with $\alpha=0.05$) for the Tversky loss and 0.15 for the Edge loss.

Let's give some more details about the training and evaluation of the models. The implementation of the methods and the hardware used are practically the same as those presented for the preliminary tests (see previous section). No regularisation or normalisation was used in the training. The models were trained for 200 epochs using the Adam optimiser, with an initial adaptive learning rate of 0.00001 and cosine decay. Our own dataset was used for training. A full 3-fold validation was not performed in our study due to time constraints. All three models were trained only once, with folds 1 and 2, and tested on fold 3. The results obtained in this way are reported in the sequel.

We evaluated the effectiveness of the methods using established metrics such as Precision, Recall, and F1 score. The methods were evaluated both in terms of their effectiveness in detecting UXO (we used a threshold of 0.5 for the overlap between prediction and ground truth) and in terms of how well each UXO was detected (undetected UXO were of course excluded from these statistics). In both cases, we will present the results in two ways: i) per-dataset statistics and ii) per-image statistics. In the first case, we will calculate the statistics for the test set as a whole (i.e. as if all the test images were combined into one large common image). In the second case, we calculate the statistics for the test set as the average of the statistics calculated for the individual images in the test set.

4.2 Quantitative and qualitative results

Firstly, the effectiveness of the UXO detection is indicated by the so-called 'per-dataset' statistics. There was a total of 151 UXOs (i.e., ground truth regions) in the entire test set, while methods M1 to M3 predicted between 211 and 295 regions, with matches between 99 and 110 regions. Table 4 summarises these statistics for all

three methods. The M3 model proved to be the best, with the highest F1 score and precision.

Table 4: Effectiveness of UXO detection: ‘per-dataset’ statistics. Best results are in bold.

Model	F1 score	Precision	Recall
M1	0,488	0,381	0,682
M2	0,493	0,373	0,728
M3	0,547	0,469	0,656

We also provide the effectiveness of UXO detection in the ‘per-image’ way, where the calculated metrics for each image are averaged over the entire test set. Such calculated metrics are shown in Table 5 (mean and standard deviation). Again, the M3 method proves to be the best. However, it should be noted that method M1 did not segment any UXO in 2 images. Methods M2 and M3 did not detect any UXO in 4 images. Table 5 also takes into account undetected UXO.

Table 5: Effectiveness of UXO detection: ‘per-image’ statistics. Best results are in bold.

Model	F1 score	Precision	Recall
M1	0.549 ± 0.305	0.515 ± 0.324	0.655 ± 0.313
M2	0.556 ± 0.289	0.508 ± 0.321	0.711 ± 0.265
M3	0.583 ± 0.285	0.582 ± 0.336	0.634 ± 0.279

We also wanted to answer the following question: If the UXO was detected, how well was it detected? Again, we calculated statistics using both ways, but since the differences are extremely small, we present only the statistics calculated using the ‘per-dataset’ way. Table 6 shows these results (mean and standard deviation). Even according to these statistics, the M3 method was the best performing method.

Table 6: Quality of UXO detection: ‘per-dataset’ statistics. Best results are in bold.

Model	F1 score	Precision	Recall
M1	0.817 ± 0.108	0.853 ± 0.140	0.813 ± 0.145
M2	0,815 ± 0.111	0,816 ± 0.144	0,850 ± 0.152
M3	0,832 ± 0.009	0,882 ± 0.117	0,813 ± 0.142

Finally, we show an example of successful and less successful UXO detection using the most promising model, M3. An example of successful detection is shown in Figure 3. The quality of UXO detection was extremely high (average F1 score was 0.945) for this example image.



Figure 3: Example of successful UXO detection with the M3 model: original RGB image (left) and model prediction overlaid on the RGB image (right).

Source: Own.

Figure 4 shows an example of unsuccessful detection with the M3 model. For this example image, the quality of the detected UXO was very low (average F1 score was 0.550), with two UXO not detected at all.



Figure 4: Example of unsuccessful UXO detection with the M3 model: original RGB image (left) and model prediction overlaid on the RGB image (right).

Source: Own.

5 Conclusion

Our preliminary research has confirmed the feasibility of UXO detection from HSI images using deep learning. With a detection rate between 60 and 70% and the F1 score above 0.8, the approach is certainly encouraging, but considering all other metrics, it is still not efficient enough for wider applicability. The main problem lies in the structure of UXO, which consists of different parts made of different materials and colours. All of this results in different spectral signatures, which were not properly accounted for in the CNN.

In further research, we will focus on improving our computational model, where we will try to account for the diversity of materials in the method. We will also investigate the implementation of a multi-class approach (i.e., including the UXO class). We will perform a full 3-fold cross-validation by testing the models.

References

- Ammunition storage area explosions – EOD clearance.* (2021). United Nations Office for Disarmament Affairs.
- Bajić, M., & Bajić, M. (2021). Modeling and Simulation of Very High Spatial Resolution UXOs and Landmines in a Hyperspectral Scene for UAV Survey. *Remote Sensing*, 13(5), 837. <https://doi.org/10.3390/rs13050837>
- Bajić, M., Ivelja, T., Krtalić, A., Tomić, M., & Vuletić, D. (2013). The multisensor and hyper spectral survey of the UXO around the exploded ammunition depot, of the land mines test site vegetation. *Proceedings 10th International Symposium HUDEM, ISSN, 9206*, 91–96.
- Bajić, M., & Potočnik, B. (2024). Spectral Response of Two Hyperspectral Cameras for UXO Endmember Selection. *ROSUS 2024 - Računalniška Obdelava Slik in Njena Uporaba v Sloveniji 2024: Zbornik 18. Strokovne Konference*, 111–120. <https://doi.org/10.18690/um.feri.1.2024.9>
- Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In V. Ferrari, M. Hebert, C. Sminchisescu, & Y. Weiss (Eds.), *Computer Vision – ECCV 2018* (Vol. 11211, pp. 833–851). Springer International Publishing. https://doi.org/10.1007/978-3-030-01234-2_49
- Guthrie, W. F. (2020). *NIST/SEMATECH e-Handbook of Statistical Methods (NIST Handbook 151)* [Dataset]. National Institute of Standards and Technology. <https://doi.org/10.18434/M32189>
- iMMAP-IHF, *Humanitarian Access Response—Monthly security incidents situation report, November 2018.* (2019). iMMAP.
- Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3431–3440. <https://doi.org/10.1109/CVPR.2015.7298965>
- National University of Public Service, Hungary, & Ember, I. (2021). The role and the risks of explosive ordnance decontamination in Hungary. *Science & Military*, 16(1), 32–42. <https://doi.org/10.52651/sam.a.2021.1.32-42>

- Roberts, S., & Williams, J. (1995). *After the guns fall silent: The enduring legacy of landmines*. Veterans of America Foundation, Washington D.C., USA.
- Ronneberger, O., Fischer, P., & Brox, T. (2015). *U-Net: Convolutional Networks for Biomedical Image Segmentation* (Version 1). arXiv. <https://doi.org/10.48550/ARXIV.1505.04597>
- Tuohy, M., Baur, J., Steinberg, G., Pirro, J., Mitchell, T., Nikulin, A., Frucci, J., & De Smet, T. S. (2023). Utilizing UAV-based hyperspectral imaging to detect surficial explosive ordnance. *The Leading Edge*, 42(2), 98–102. <https://doi.org/10.1190/tle42020098.1>
- Yan, K., Wang, X., Lu, L., Zhang, L., Harrison, A. P., Bagheri, M., & Summers, R. M. (2018). Deep Lesion Graphs in the Wild: Relationship Learning and Organization of Significant Radiology Image Findings in a Diverse Large-Scale Lesion Database. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9261–9270. <https://doi.org/10.1109/CVPR.2018.00965>

ANALIZA ODBOJKARSKE IGRE Z UPORABO ALGORITMOV RAČUNALNIŠKEGA VIDA

MARKO PLANKELJ, UROŠ MLAKAR

Univerza v Mariboru, Fakulteta za elektrotehniko računalništvo in informatiko, Maribor, Slovenija

marko.plankelj@student.um.si, uros.mlakar@um.si

V zadnjih letih so sodobne tehnologije naredile šport bolj dostopen širšemu občinstvu z zagotavljanjem interaktivnih podatkov med prenosi, zmanjšanjem tveganja človeške napake in izboljšanjem uspešnosti športnikov s pomočjo realno časovne analize in ciljnih vpogledov v trening. Ta članek združuje teoretične in praktične pristope z razvojem aplikacije, ki temelji na specifičnih konvolucijskih nevronskih mrežah za zaznavanje igrišča za odbojko in sledenje žogi. Rezultati prikazujejo sposobnost napredne video analitike v športu, ki uporabnikom omogoča raziskovanje priložnosti sodobne tehnologije pri izboljšanju športne uspešnosti.

DOI
[https://doi.org/
10.18690/um.feri.2.2025.8](https://doi.org/10.18690/um.feri.2.2025.8)

ISBN
978-961-286-960-1

Ključne besede:
računalniški vid,
konvolucijske nevronske
mreže,
detekcija objektov,
odbojka,
spletna aplikacija

Prispevek temelji na:
Plankelj, M. (2025). *Analiza odbojcarske igre z uporabo algoritmov računalniškega vida in strojnega učenja*. magistrsko delo, Univerza v Mariboru, Fakulteta za elektrotehniko, računalništvo in informatiko. Maribor.



Univerzitetna založba
Univerze v Mariboru

DOI
[https://doi.org/
10.18690/um.feri.2.2025.8](https://doi.org/10.18690/um.feri.2.2025.8)

ISBN
978-961-286-960-1

Keywords:

computer vision,
convolutional neural
networks,
object detection,
volleyball,
web application

The proceedings is based

on: Plankelj, M. (2025).
*Analiza odbojske igre z
uporabo algoritmov
računalniškega vida in strojnega
učenja*: magistrsko delo,
University of Maribor,
Faculty of Electrical
Engineering and Computer
Science. Maribor

VOLLEYBALL GAME ANALYSIS USING COMPUTER VISION ALGORITHMS

MARKO PLANKELJ, UROŠ MLAKAR

University of Maribor, Faculty of Electrical Engineering and Computer Science,
Maribor, Slovenia
marko.plankelj@student.um.si, uros.mlakar@um.si

In recent years, modern technologies have made sports more accessible to a wider audience by providing interactive data during broadcasts, reducing the risk of human error, and enhancing athletes performance through real-time analysis and targeted training insights. This paper combines theoretical and practical approaches by developing an application based on specific convolutional neural networks for volleyball court detection and ball tracking. The results demonstrate the potential of advanced video analytics in sports, allowing users to explore the opportunities of modern technology in improving sports performance.



1 Uvod

Šport je za ljudi že od nekdanj oblika sprostitve, druženja, priložnost za iskanje novih prijateljstev z dokazano koristnimi učinki na zdravje. Svetovna zdravstvena organizacija priporoča telesno aktivnost kot del zdravega načina življenja v programu, s pomočjo katerega želijo ljudem približati telesno aktivnost kot priložnost za bolj zdravo, srečno in produktivno življenje (World Health Organization, 2024). Na drugi strani iz dneva v dan spremljamo razvoj sodobnih tehnologij in njihovo uporabo na najrazličnejših področjih, med drugim tudi v športu.

Šport in sodobne tehnologije, kot sta računalniški vid in strojno učenje, sta bila, še pred nekaj leti, povsem nezdržljiva pojma, danes pa si skorajda ne moremo predstavljati spremljanja ali udeleževanja v športu brez uporabe sodobnih tehnologij (B.T. Naik, 2022). Kljub številnim izzivom, kot so slabša kvaliteta vhodnih podatkov ali prekrivanja igralcev, je uporaba sistemov za detekcijo igrišča, igralcev in njihovih akcij ter sledenja žogi med igro vse pogostejša. Z njihovo uporabo želijo preprečiti sporne situacije na tekmah, izboljšati treninge in analizo tekme, z namenom preprečevanja poškodb predvidevati obremenitve na treningih in tekmah ter izboljšati izkušnjo gledalcem z analizo pred, med in po tekmi (Plankelj, 2025).

2 Predstavitev problema

Obravnavan problem in predstavljeno rešitev lahko razdelimo na štiri korake:

1. Priprava podatkov za učenje, ki je vključevala zbiranje, predobdelavo in označevanje podatkov, ki smo jih kasneje dodatno augmentirali z namenom povečanja raznolikosti in obsega nabora podatkov.
2. Implementacija in učenje modelov konvolucijske nevronske mreže na podlagi pripravljenega nabora podatkov.
3. Perspektivna transformacija odbojcarskega igrišča.
4. Razvoj spletne aplikacije in integracija naučenih modelov konvolucijske nevronske mreže v povezavi s perspektivno projekcijo za prikaz končnih rezultatov.

2.1 Obstoječe rešitve

Uporaba sodobnih tehnologij ni več omejena na pilotne projekte in dogodke nižjega ranga, temveč se vse bolj uveljavlja na najvišjih svetovnih športnih dogodkih.

Na lanskih olimpijskih igrah v Parizu so v sodelovanju s pomočjo partnerja, podjetja Intel, predstavili tehnologije za izboljšanje izkušenj udeležencev in gledalcev. Mednarodni olimpijski komite je med drugim tudi oznanil začetek celostnega programa uporabe umetne inteligence v športu, s katerim želijo nasloviti področja vključevanja umetne inteligence v šport, in sicer z namenom izboljšanja tako uspešnosti športnikov kot izkušnje gledalcev (Olympics, 2024).

Na svetovnem nogometnem prvenstvu v Katarju leta 2022 pa so po več uspešnih testiranjih za preverbo prepovedanega položaja med igro vpeljali tudi uporabo polavtomatske tehnologije. Tehnologija uporablja dvanajst kamer, nameščenih pod vrhom stadiona, za izračun položaja devetindvajsetih ključnih točk na vsakem igralcu petdesetkrat na sekundo. Za natančno zaznavanje udarca žoge uporabljajo senzor IMU (angl. inertial measurement unit), ki se nahaja na sredini žoge in posreduje podatke o žogi v sobo za pregled posnetka petstokrat na sekundo (Inside FIFA, 2024).

V športih, kot sta tenis in odbojka, se že vrsto let uporablja sistem Hawk-eye za sledenje poti žoge in določanje njenega položaja s pomočjo hitrih kamer, nameščenih okrog igralne površine. Sistem v vsakem izmed njih identificira piksle, ki ustrezajo žogi, in nato s pomočjo vsaj dveh slik, posnetih iz drugih kamer (ki so postavljene na druge lokacije), primerja njen položaj in ga potrди oz. primerno popravi (Hawk Eye, 2024).

Kot že omenjeno, tudi v odbojki, tako kot v drugih športih, uvajanje naprednih tehnologij ni izjema. Platforma Balltime z umetno inteligenco Volleyball AI (VOLL-E) razdeli odbojgarsko igro v različne segmente, s pomočjo katerih olajša analizo tekme in pripravo igralcev na naslednje tekme. Z uporabo modela konvolucijske nevronske mreže, ki se je učil na obsežni zbirki odbojgarskih posnetkov, se nato iz naloženega videoposnetka tekme omogoča avtomatska zaznava žoge ter vsakega izmed igralcev na igrišču. Na podlagi prepoznanih položajev žoge in igralcev platforma prepozna akcije, kot sta sprejem ali obramba, in samodejno določi smer napada ter jo vizualno predstavi. S pomočjo zbranih podatkov izračuna hitrost žoge

in sortiranje ter izbiro elementov igre glede na posameznika v igrišču (Balltime Academy, 2024).

Podobno funkcionalnost, kot jo ponuja zgoraj opisana platforma Balltime, omogoča tudi aplikacija za mobilne naprave znamke Apple, imenovana Avais, ki za razliko spremlja in analizira odbojcarsko igro v realnem času. Posledično se izognemo čakanju med nalaganjem videoposnetka odbojcarske tekme in lahko podatke za analizo pridobimo še v istem trenutku (Avais, 2024).

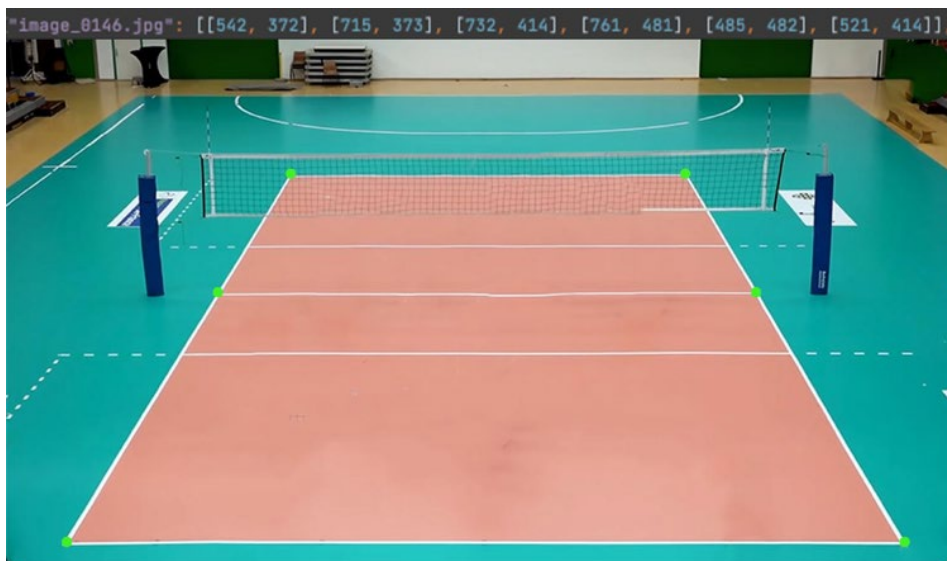
3 Implementacija lastne rešitve

Končna rešitev je bila izvedena v več korakih z uporabo različnih tehnologij. Z namenom boljše in lažje dostopnosti, neodvisno od lokacije in naprave končnega uporabnika, je bila razvita spletna aplikacija, ki uporabi naučene modele nevronske mreže za zaznavo odbojcarskega igrišča ter sledenje odbojcarski žogi skozi slike v krajšem izseku.

3.1 Zbiranje, predobdelava, označevanje ter augmentacija učnih podatkov

Primarno zbiranje podatkov je potekalo s pomočjo prosto dostopnih podatkov na svetovnem spletu. Zaradi zamudnega iskanja podatkov, ki zadoščajo postavljenim merilom (slika celotnega odbojcarskega igrišča, zajetega s pomočjo ene kamere, postavljene za odbojcarskim igriščem) smo kot drugi vir podatkov zajeli lastni posnetek odbojcarske tekme. Za pretvorbo iz video sekvence v slike smo implementirali skripto v programskem jeziku Python in jih shranili v formatu JPG (angl. Joint Photographic Experts Group) v zaporednem časovnem intervalu ene slike na sekundo.

Zaradi različnih virov podatkov smo v predobdelavi podatkov najprej želeli poenotiti dimenzije vseh slik, za kar smo pripravili enostavno skripto in poenotili dimenzije vseh slik spremenili na želeno velikost. Nato smo nadaljevali z označevanjem učnih podatkov z uporabo dveh ločenih metod. Pri prvi metodi smo ročno izbrali šest točk na vsaki sliki in nato za vsako od šestih izbranih točk shranili koordinate x in y v formatu JSON (angl. JavaScript Object Notation) za kasnejšo uporabo, kot je prikazano na sliki 1.



Slika 1: Program, ki omogoča označbo zelenih ključnih točk na odbojgarskem igrišču.

Vir: lasten.

Pri drugi tehniki označevanja podatkov, uporabljeni v učni množici za zaznavanje žog, smo izkoristili funkcionalnost spletne platforme Roboflow, ki razvijalcem ponuja celovite storitve za izdelavo aplikacij računalniškega vida, vključno z označevanjem podatkov v učni množici (Roboflow, 2024).

Zaradi manjšega števila podatkov v učnem naboru smo se odločili za augmentacijo učnih podatkov, kot je prikazano na sliki 2. Uporabili smo **imgaug**, namensko knjižnico za povečanje učne množice s pomočjo različnih tehnik augmentacije, ki jih podpira (Imgaug, 2024). Za ohranitev celotnega igrišča na sliki po augmentaciji smo prebrali koordinate kotov odbojgarskega igrišča in glede na njihovo oddaljenost od roba slike določili transformacije, ki smo jih uporabili. Transformacije so bile izvedene v naključnem vrstnem redu s funkcijo Sequential. Po transformacijah smo preverili, ali so vse označene točke igrišča ostale znotraj slike. V primeru, da je katera koli izmed točk padla izven omejenega območja, smo postopek transformacije ponovili največ petkrat in v primeru neuspeha s polovično verjetnostjo aplicirali zgolj horizontalno zrcaljenje.



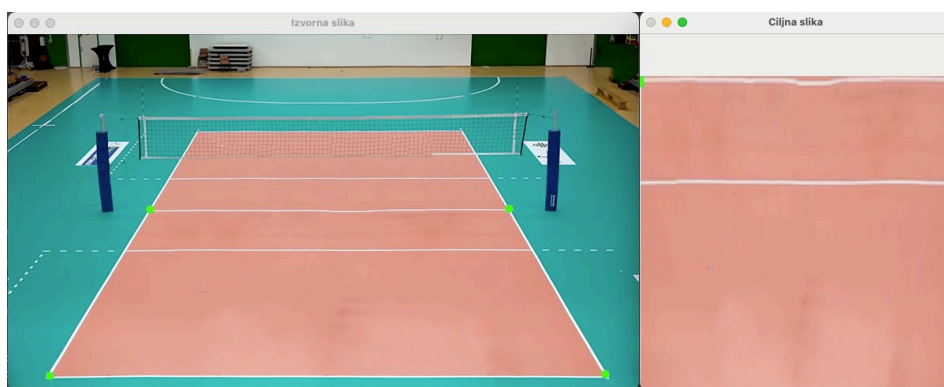
Slika 2: Praktičen prikaz podatkov pred in po augmentaciji (afina transformacija).

Vir: lasten.

3.2 Perspektivna transformacija odbojkarskega igrišča

Perspektivne transformacije se uporabljajo na različnih področjih, vključno z avtonomno vožnjo. Pri tem se posnetki več kamer, nameščenih na vozilo, s pomočjo perspektivne transformacije pretvorijo v prikaz iz ptičje perspektive, ki zajema celotno okolico vozila. To omogoča lažje ocenjevanje razdalj med objekti v okolici (Joseph Redmon, 2016).

V našem primeru smo za izvedbo perspektivne transformacije uporabili knjižnico OpenCV, specializirano za računalniški vid. Najprej smo izračunali matriko homografije in jo nato uporabili za preoblikovanje med izvornimi in ciljnimi točkami. Končni rezultat je pogled iz ptičje perspektive na odbojkarsko igrišče, kot je prikazano na sliki 3, kar omogoča enostavnejšo analizo igre.



Slika 3: Prikaz perspektivne transformacije slike.

Vir: lasten.

3.3 Implementacija in učenje modelov konvolucijske nevronske mreže

Za učenje smo izbrali dva ločena modela konvolucijske nevronske mreže: segmentacijsko nevronske mrežo U-Net, ki je bila uporabljena za zaznavo odbojarskega igrišča, ter model YOLOv8, ki je bil uporabljen za zaznavo žoge. V obeh primerih smo temeljili na pristopu, kjer smo za učenje in testiranje uporabljali ločene učne množice, torej med učenjem nismo uporabili nobene izmed slik, ki bi bila kasneje uporabljena za testiranje uspešnosti katerega izmed modelov nevronskih mrež.

Segmentacijsko nevronske mrežo U-Net, katere simetrično strukturo sestavljata kodirnik in dekodirnik, je bila implementirana z uporabo odprtokodnega ogrodja za strojno učenje, imenovan PyTorch.

Za model YOLO smo se odločili uporabiti eno od novejših verzij, natančneje verzijo osem, razvito s strani podjetja Ultralytics (Ultralytics, 2024). Čeprav bi lahko učenje izvajali na platformi podjetja Ultralytics, smo knjižico Ultralytics raje namestili lokalno ter nato integrirali v Python program. Za namen učenja smo uporabili predhodno učen model YOLOv8, ki smo ga s pomočjo predhodno označenih podatkov nato zgolj dodatno učili na lastnih podatkih. Lastne podatke smo imeli shranjene v formatu za serializacijo, imenovanem YAML, ki smo ga predhodno generirali na platformi Roboflow, ob označevanju podatkov

3.4 Implementacija spletne aplikacije

Z namenom boljše in lažje dostopnosti, neodvisno od lokacije in naprave končnega uporabnika, smo razvili spletno aplikacijo, katere glavni namen je uporaba naučenih modelov nevronskih mrež za zaznavo odbojarskega igrišča ter sledenje odbojarski žogi skozi slike v krajšem izseku odbojarske tekme. Razvoj je potekal v programskem okolju PyCharm, kjer smo za zaledje aplikacije uporabili spletno mikro-ogrodje Flask, medtem ko smo videz nadgradili in izboljšali z odprtokodnim CSS ogrodjem Bootstrap.

4 Rezultati

V tem poglavju so predstavljeni tako rezultati implementirane aplikacije kakor tudi glavne funkcionalnosti le te.

4.1 Predstavitev aplikacije in uporabniškega vmesnika

Aplikacija je idejno razdeljena na dva glavna dela. V prvem delu lahko uporabnik izbira med predpripravljenimi posnetki ali izbere lasten videoposnetek iz naprave, preko katere dostopa do spletne aplikacije. Ob uspešni izbiri, lahko uporabnik začne analizo katere rezultat se prikaže v štirih predpripravljenih poljih (kot je prikazano na sliki 4):

- **Zaznava igrišča** – zaznava odbojcarskega igrišča na prvi sliki videoposnetka s pomočjo naučenega modela konvolucijske nevronske mreže z arhitekturo U-Net. Kote odbojcarskega igrišča označimo z rdečimi pikami.
- **Sledenje žogi** – zaznava odbojcarske žoge na prvi sliki videoposnetka s pomočjo naučenega modela konvolucijske nevronske mreže z arhitekturo YOLO. Zaznano odbojcarsko žogo označimo s piko zelene barve.
- **Homografija igrišča** – perspektivna transformacija odbojcarskega igrišča, s pomočjo katere za lažjo analizo poti žoge ustvarimo pogled od zgoraj navzdol.
- **Smer napada** – prikaz gibanja odbojcarske žoge skozi zaporedje slik celotne dolžine videoposnetka, kjer za vsako sliko shranimo podatke o lokaciji žoge in ob koncu povežemo zaporedne položaje ter jih prikažemo kot pot njenega premikanja na igrišču, prikazanem s ptičje perspektive. Pot žoge označimo s črto rdeče barve.



Slika 4: Rezultati analize izbranega videoposnetka.

Vir: lasten.

4.2 Statistična in vizualna primerjava rezultatov

Predstavitvi aplikacije in uporabniškega vmesnika sledijo rezultati uspešnosti zaznave odbojarskega igrišča in žoge, kar smo testirali na različnih testnih podatkih.

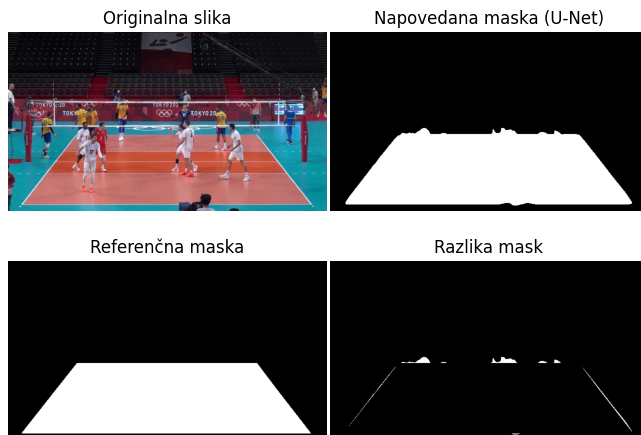
4.2.1 Analiza zaznave odbojarskega igrišča

Rezultate zaznave igrišča smo statistično ovrednotili s pomočjo standardne metrike, ki se uporablja pri segmentaciji, in sicer oceno prekrivanja napovedanih segmentov IoU (angl. Intersection over Union), ki meri natančnost prekrivanja med napovedanimi segmenti (A) in dejanskimi ročnimi oznakami (B) po naslednji formuli:

$$IoU = \frac{|A \cap B|}{|A \cup B|} \quad (1)$$

Model smo testirali na naboru 174 naključno izbranih slik, ki so bile posnete z različnih zornih kotov, v različnih športnih dvoranah, s čimer smo omogočili oceno robustnosti in prilagodljivosti modela. Model je dosegel povprečno vrednost 0,8627, kar pomeni, da se v povprečju 86,27 % segmentiranih površin ujema z ročno označenimi segmenti odbojarskega igrišča. Visok odstotek potrjuje uspešnost modela pri zaznavi odbojarskega igrišča, kljub temu pa smo opazili določene primere in pogoje, pri katerih model ne vrne pričakovanih rezultatov.

Slika 5 prikazuje primer, kjer so robovi (natančneje zgornji rob) odbojarskega igrišča prekriti z igralci. V tem primeru je model imel težave pri zaznavanju segmentov igrišča, saj so igralci blokirali vidne meje in s tem zameglili robove igrišča. Posledično so bile napovedane meje manj ostre in natančne, kar je pripeljalo do višje stopnje napak pri segmentaciji (in posledično nižje vrednosti metrike IoU). Podobne težave smo opazili ob nepravilni postavitvi kamere (npr. kamera, postavljena ob stran odbojarskega igrišča).



Slika 5: Slabša segmentacija odbojkarskega igrišča ob prikritih robovih.

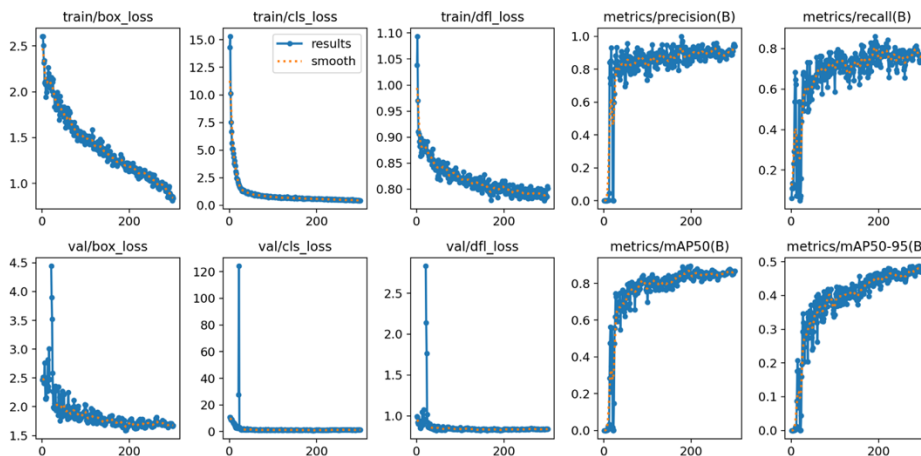
Vir: lasten.

4.2.2 Analiza zaznave odbojkarske žoge

Po zaključku učenja smo analizirali grafe, ki prikazujejo rezultate na učni in validacijski množici. Omogočajo nam pomemben vpogled v delovanje modela in ocene njegove generalizacije in robustnosti. V analizi se osredotočimo zgolj na nekaj izmed ključnih metrik oziroma grafov, ki jih prikazuje slika 6:

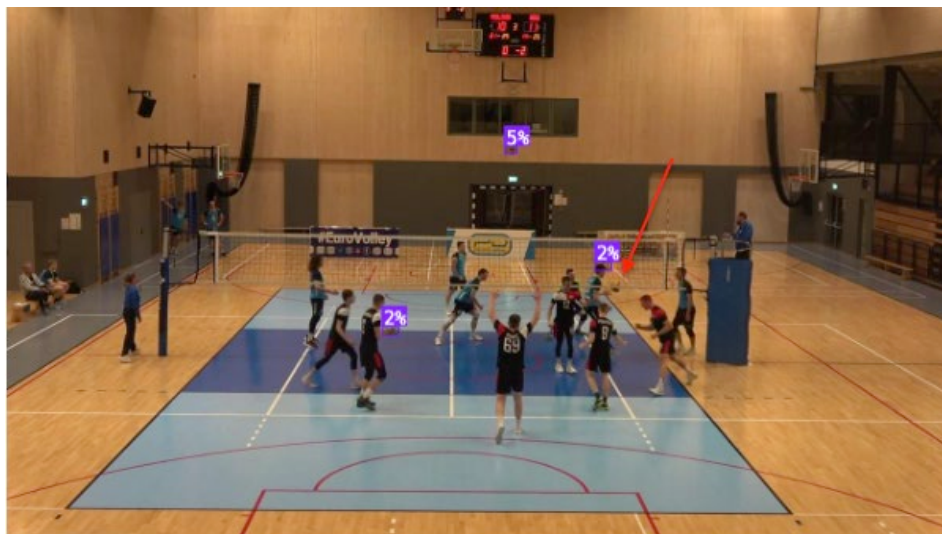
- Train/box_loss in val/box_loss – metrika, ki predstavlja natančnost zaznanih okvirjev glede na dejanski položaj žoge. Padajoči graf predstavlja izboljšanje učenja in bolj natančno zaznavo žoge, hkrati pa opazimo zmanjšanje izgube na validacijski množici. Pri izgubi na validacijski množici smo dosegli vrednost okoli 1.5, kar kaže na ustrezno generalizacijo modela tudi na nove podatke (slike), ki niso bili del učne množice, brez očitnih znakov prenaučnosti.
- Metrics/mA50(B) – metrika, ki predstavlja povprečno natančnost zaznanih okvirjev pri 50 % prekrivanju med napovedanim in dejanskim okvirjem, torej 50 % vrednostjo IoU (glej poglavje 4.2.1). Vrednost, ki jo dosega naš model, je 85 %, kar kaže na visoko stopnjo natančnosti pri nižjem pragu prekrivanja.
- Metrics/mAP50-95(B) – pri višjem pragu prekrivanja (povprečna natančnost preko različnih pragov prekrivanja od 50 do 95 %), ki je prikazan

na grafu z naslovom `metrics/mAP50-95(B)`, je imel model večje težave pri natančnem umeščanju okvirjev, zlasti v situacijah, kjer se žoga nahaja blizu drugih objektov ali v slabših svetlobnih pogojih.



Slika 6: Grafični prikaz rezultatov učenja..

Vir: lasten.



Slika 7: Neuspešna zaznava odbojcarske žoge ob delnem prekrivanju žoge.

Vir: lasten.

Za celovito oceno delovanja modela smo rezultate preverili tudi vizualno, s prikazom zaznanih pozicij odbojcarske žoge na slikah, ki niso bile del učne množice. Testiranje modela smo lahko izvedli na platformi Roboflow. Rezultati so bili uspešni v primeru, da je žoga na sliki jasno vidna, na sliki ni prisotnih več žog, osvetlitveni pogoji pa so optimalni. Neuspešna zaznava se je pojavila v primeru, ko pogoji niso bili idealni, kot prikazuje slika 7. Žoga na sliki ni jasno vidna zaradi prekrivanja z drugimi objekti (odbojcarsko mrežo), zaradi česar je imel model težave pri zaznavi žoge in le to zaznal na več mestih zaznal potencialno prisotnost odbojcarske žoge, vendar z nizkimi stopnjami zaupanja. Rdeča puščica na sliki 7 označuje dejansko lokacijo žoge.

Kljub prisotnosti motečih elementov na igrišču in v ozadju je model sposoben uspešno prepoznati pravo žogo, kot je prikazano na sliki 8, kjer slika vsebuje več kot eno odbojcarsko žogo. Model je pravilno zaznal lokacijo obeh žog z različnima stopnjama zaupanja, pri čemer je prava lokacija žoge označena z višjo, 88-odstotno stopnjo zaupanja.



Slika 8: Uspešna zaznava in izbira odbojcarske žoge.

Vir: lasten.

4.3 Možne izboljšave

Kljub uspešni implementaciji zelenih funkcionalnosti se aplikacija v določenih situacijah ne obnese odlično. To postane očitno predvsem v primerih, ko vhodni video vsebuje slike, ki se razlikujejo od tistih, ki se uporabljajo za usposabljanje konvolucijskih nevronske mreže. Najpogostejše težave, ki smo jih opazili, so bile:

- Različni položaji kamere: ko je odbojcarska tekma posneta pod kotom kamere, ki se razlikuje od tistih, uporabljenih v naboru podatkov o treningu.
- Več vrstic na igrišču: ko videoposnetek vključuje več vrstic, ki se ne nanašajo samo na igrišče za odbojko.
- Barvna shema žoge: v nekaterih ligah (tudi v najvišjih ligah, na primer v Italiji) uporabljajo žogo druge barve. Poleg tega se lahko barvne sheme igrišča prekrivajo z žogo, zaradi česar je težko natančno zaznati žogo.
- Ključne točke igrišča ali žoga, ki jo pokrivajo igralci: ko ključne točke igrišča pokrivajo igralci ali druge ovire (npr. mreža, ki pokriva črto na najbolj oddaljenem igrišču od kamere, če je kamera postavljena prenizko), kar povzroči težave pri zaznavanju igrišča za odbojko ali žoge.
- Omejitve obdelave v realnem času: glede na računalniške vire in kompleksnost spletne aplikacije lahko pretakanje v realnem času in analiza videa (zlasti z višjo ločljivostjo) povzročita zakasnitev ali vplivata na zmogljivost.

5 Zaključek

V prispevku smo predstavili proces od začetne ideje do funkcionalne spletne aplikacije, ki nudi rešitev za prvotni koncept, ki je bil samodejna analiza in vizualna predstavitev rezultatov uporabniku. Uporabljene metode so temeljile na zbiranju, pripravi, označevanju in augmentaciji podatkov, ki so bili nato uporabljeni za učenje dveh konvolucijskih nevronske mreže. Naučeni modeli so bili nato uporabljeni v povezavi s perspektivno transformacijo igrišča za analizo in vizualno predstavitev rezultatov uporabniku. Končni rezultat je bil predstavljen kot uporabniku prijazna spletna aplikacija, kjer lahko uporabnik izbere želene video in v nekaj sekundah prejme osnovno analizo, vključno z zaznavo igrišča in sledenjem odbojcarski žogi.

Aplikacija ima kljub uspešni implementaciji nekaj slabosti in scenarijev, kjer rezultati niso takšni, kot so pričakovani. Izzivi vključujejo različne položaje kamere, barvne sheme žog in igrišč ali, motnje, ki lahko zakrijejo ključne točke igrišča ali samo žogo. Poleg tega lahko na obdelavo v realnem času vplivajo računalniške omejitve in kakovost videa, kar lahko vpliva na zmogljivost.

Kljub temu implementirana aplikacija služi kot temelj, ki omogoča številne nadgradnje, ki bi se lahko zgledovale po obstoječih rešitvah in izboljšale njihove pomanjkljivosti. Kot končni rezultat bi lahko uporabnikom posredovali realno časovno statistiko odbojcarske tekme, platformo pa bi lahko nadgradili tudi na druge športe in s tem pritegnili širši krog uporabnikov. Vsi našteti razlogi spodbujajo zavedanje, da vpeljava sodobnih tehnologij v vse segmente našega življenja, tudi šport, ni več binarno vprašanje, ampak zgolj vprašanje časa.

Viri in literatura

- Avais*. (2. 6 2024). Pridobljeno iz <https://www.avais.ai/features>
- B.T. Naik, M. H. (2022). A Comprehensive Review of Computer Vision in Sports: Open Issues, Future Trends and Research Directions. *Applied Sciences*.
- Balltime Academy*. (2. 6 2024). Pridobljeno iz What is Volleyball AI: <https://academy.balltime.com/getting-started/what-is-volleyball-ai>
- Hawk Eye*. (2. 6 2024). Pridobljeno iz <https://en.wikipedia.org/wiki/Hawk-Eye>
- Imgaug*. (2. 6 2024). Pridobljeno iz <https://imgaug.readthedocs.io/en/latest>
- Inside FIFA*. (2. 6 2024). Pridobljeno iz Semi-automated Offside Technology to be Used at FIFA World Cup 2022: <https://inside.fifa.com/technical/media-releases/semi-automated-offside-technology-to-be-used-at-fifa-world-cup-2022-tm>
- Joseph Redmon, S. D. (2016). You Only Look Once: Unified, Real-Time Object Detection. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (str. 779-788).
- Olympics*. (2. 6 2024). Pridobljeno iz IOC Takes the Lead for the Olympic Movement and Launches Olympic AI Agenda: <https://olympics.com/ioc/news/ioc-takes-the-lead-for-the-olympic-movement-and-launches-olympic-ai-agenda>
- Plankelj, M. (2025). Analiza odbojcarske igre z uporabo algoritmov računalniškega vida in strojnega učenja: magistrsko delo, Univerza v Mariboru, Fakulteta za elektrotehniko, računalništvo in informatiko. Maribor.
- Roboflow*. (2. 6 2024). Pridobljeno iz Our Company: <https://roboflow.com/about>
- Ultralytics*. (2024). Pridobljeno iz YOLOv8 Models Documentation.
- World Health Organization*. (30. 8 2024). Pridobljeno iz Sports and Health Initiative: <https://www.who.int/initiatives/sports-and-health>

Tukaj smo mi. Slovenski podjetniški sklad.

Da vam svetujemo, pomagamo, da vas podpiramo.
Da vam prisluhujemo in da vas slišimo.

podjetniskisklad.si

Slovenski podjetniški sklad (SPS) je osrednja finančna institucija, ki podpira rast in razvoj mikro, malih in srednje velikih podjetij (MSP) v Sloveniji. Deluje pod okriljem Ministrstva za gospodarstvo, turizem in šport ter podjetjem omogoča lažji dostop do financiranja s pomočjo različnih razvojnih spodbud. Cilj teh spodbud je povečati inovativnost, konkurenčnost in trajnostno rast slovenskih podjetij. V letu 2025 bo SPS namenil nekje **186 milijonov evrov** za različne razvojne spodbude, s katerimi bo podprl okoli **4.000 podjetniških projektov**.



Glavne naloge in instrumenti SPS:

- 1. Finančne spodbude** za MSP-je, start-up in scale-up podjetja s poudarkom na prioriteth razvojnih področjih
 - **Garancije za bančna posojila** s subvencijo obrestne mere, ki podjetjem omogočajo ugodnejša posojila.
 - **Mikrokrediti** z nizkimi obrestnimi merami in enostavnejšimi pogoji za mala podjetja.
 - **Subvencije in nepovratna sredstva** za inovacije, digitalizacijo, trajnostni razvoj in internacionalizacijo.
 - **Tvegani kapital in lastniško financiranje** prek različnih skladov tveganega kapitala.
- 2. Specifične spodbude** za dvig poslovnih in razvojnih kompetenc
 - **Vavčerji** – enostavne in hitro dostopne spodbude malih vrednosti
 - **Mentorski program** – več ko 70 izkušenih mentorjev in strokovnjakov

- **Programi usposabljanja in povezovanja**

3. Partnerstva za dvig inovativne usmerjenosti in trajnosti

SPS aktivno spodbuja razvoj partnerstev za povečanje inovativnosti in trajnosti med mikro, malimi in srednje velikimi podjetji (MSP), start-upi in scale-up podjetji. Ta partnerstva se osredotočajo na osem ključnih vertikalnih razvojnih področij:

- Pametna infrastruktura
- Napredne informacijsko-komunikacijske tehnologije
- Trajnostni viri in surovine
- Trajnostna hrana in turizem
- Zdravje – medicina
- Mobilnost
- Napredni materiali kot končni produkt
- Tovarne prihodnosti

Ta področja so horizontalno povezana s štirimi trajnostnimi vidiki: krožnim, zelenim, digitalnim in družbeno koristnim. SPS povezuje podjetja z institucijami znanja, raziskovalnimi organizacijami in podpornimi institucijami, najprej na nacionalni ravni, nato pa tudi globalno.

4. Promocijske aktivnosti za krepitev podjetništva in inovativnosti

SPS izvaja številne promocijske aktivnosti s katerimi spodbuja podjetništvo, inovativnost in trajnostni razvoj. Namen teh aktivnosti je ozaveščanje, povezovanje ter podpora mikro, malim in srednje velikim podjetjem (MSP), start-upom in scale-up podjetjem. Med ključne promocijske aktivnosti spadajo: organizacija dogodkov, delavnic in konferenc, sodelovanje s podjetniškimi podpornimi mrežami, promocija inovativnih in trajnostnih projektov itd...

5. Učinkovita javna finančna institucija z upoštevanje ESG načel

Ciljne skupine

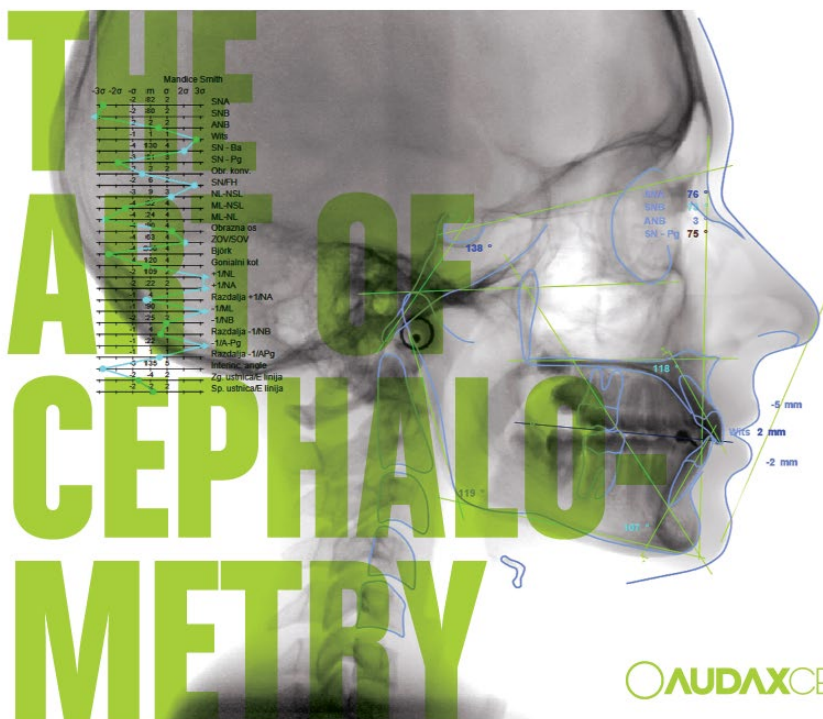
- Start-up in scale-up podjetja, ki potrebujejo začetni kapital za rast in razvoj.
- Mikro, mala in srednje velika podjetja, ki želijo vlagati v inovacije, digitalizacijo in rast.
- Podjetja, ki iščejo ugodne vire financiranja za širitev na tuje trge.

Pomen SPS za slovensko gospodarstvo

Slovenski podjetniški sklad igra ključno vlogo pri spodbujanju podjetništva in gospodarskega razvoja, saj zmanjšuje finančne ovire za podjetja ter povečuje njihovo konkurenčnost na domačem in mednarodnem trgu.

Za več informacij o aktualnih razpisih in pogojih financiranja obiščite [uradno spletno stran SPS](#).





AudaxCeph. A new paradigm of computer-based cephalometric analyses.

Use only the best-of-breed tools in your work. AudaxCeph persuades with its high-end performance, numerous features and ease of use.

The AudaxCeph software suite includes tools which are required for creating cephalometric analyses, superimposition and planning orthodontic treatment. It provides flexible and easy-to-use tools to quickly and efficiently process even the most difficult cases in orthodontics, dentistry and maxillofacial surgery. AudaxCeph offers an interface in many languages and can be easily translated to any language using Latin or Cyrillic alphabet. Technical support is available and can be provided on-line.

AUDAXCEPH®

www.audaxceph.com
info@audaxceph.com
 AUDAX (fax) - 3862604000 - 3-810104000
 tel. +386 120140101 - fax. +386 14247101

ROSUS 2025 - RAČUNALNIŠKA OBDELAVA SLIK IN NJENA UPORABA V SLOVENIJI 2025: ZBORNİK 19. STROKOVNE KONFERENCE

BOŽIDAR POTOČNIK (UR.)

Univerza v Mariboru, Fakulteta za elektrotehniko, računalništvo in informatiko, Inštitut
za računalništvo, Maribor, Slovenija
bozidar.potocnik@um.si

ROSUS 2025 – Računalniška obdelava slik in njena uporaba v Sloveniji 2025 je strokovna računalniška konferenca, ki jo od leta 2006 naprej vsako leto organizira Inštitut za računalništvo iz Fakultete za elektrotehniko, računalništvo in informatiko, Univerze v Mariboru. Konferenca povezuje strokovnjake in raziskovalce s področij digitalne obdelave slik in strojnega vida z uporabniki tega znanja, pri čemer uporabniki prihajajo iz raznovrstnih industrijskih okolij, biomedicine, športa, zabavnštva in sorodnih področij. Zbornik konference ROSUS 2025 združuje strokovne prispevke več avtorjev, od tega dve vabljeni predavanji ter več demonstracijskih prispevkov. Prispevki podajajo najnovejše dosežke slovenskih strokovnjakov s področij digitalne obdelave slik in strojnega vida, osvetljujejo pa tudi trende in novosti na omenjenih strokovnih področjih. Velik poudarek prispevkov je na promoviranju ekonomske koristnosti aplikacij računalniške obdelave slik in vida v slovenskem prostoru. Takšne računalniške aplikacije zaradi visoke natančnosti, robustnosti in izjemnih hitrosti pri obdelovanju informacij nudijo namreč nove priložnosti za uveljavitev na trgu visokih tehnologij.

DOI
[https://doi.org/
10.18690/um.feri.2.2025](https://doi.org/10.18690/um.feri.2.2025)

ISBN
978-961-286-960-1

Ključne besede:
računalniška obdelava slik,
strojni vid,
biomedicina,
industrijske aplikacije,
prenos znanja



Univerzitetna založba
Univerze v Mariboru

DOI
[https://doi.org/
10.18690/um.feri.2.2025](https://doi.org/10.18690/um.feri.2.2025)

ISBN
978-961-286-960-1

ROSUS 2025 – COMPUTER IMAGE PROCESSING AND ITS APPLICATION IN SLOVENIA 2025: PROCEEDINGS OF THE 19TH PROFESSIONAL CONFERENCE

BOŽIDAR POTOČNIK (ED.)

University of Maribor, Faculty of Electrical Engineering and Computer Science, Institute
of Computer Science, Maribor, Slovenia
bozidar.potocnik@um.si

Keywords:

evolutionary algorithm,
deep networks,
image classification,
generating deep networks,
optimization

ROSUS 2025 – Computer image processing and its application in Slovenia 2025 is a professional conference that, since 2006, has been organised every year by the Institute of Computer Science of the Faculty of Electrical Engineering and Computer Science, University of Maribor. The conference connects researchers in the fields of Image Processing and Machine Vision with users of this knowledge, whereby users are coming from diverse industrial environments, such as Biomedicine, Sport, Entertainment, and related fields. The proceedings of ROSUS 2025 combine scientific articles by dozens of authors, including two invited lectures and several demonstration articles. Contributions represent the latest achievements of Slovenian experts in the fields of Image Processing and Vision, and also highlight trends and novelties in these areas. Great emphasis is on promotion of the economic usefulness of Image Processing and Vision applications in the Slovenian region. Namely, such software, due to high precision, robustness, and exceptional speed in information processing, provides new opportunities for penetration on the high technologies market.



University of Maribor Press



Univerza v Mariboru

Fakulteta za elektrotehniko,
računalništvo in informatiko

LSPO

Laboratorij za sistemsko
programsko opremo

<https://rosus.feri.um.si/rosus2025>
rosus.feri@um.si

