

Zbornik strokovne konference

ROSUS 2026

Računalniška obdelava slik in
njena uporaba v Sloveniji 2026

Urednik:
Božidar Potočnik



Maribor, 19. marec 2026



ROSUS 2026 – Računalniška obdelava slik in njena uporaba v Sloveniji 2026

Zbornik 20. strokovne konference

Urednik
Božidar Potočnik

Marec 2026

Naslov <i>Title</i>	ROSUS 2026 - Računalniška obdelava slik in njena uporaba v Sloveniji 2026 <i>ROSUS 2026 - Computer image processing and its application in Slovenia 2026</i>
Podnaslov <i>Subtitle</i>	Zbornik 20. strokovne konference <i>Proceedings of the 20th Professional Conference</i>
Urednik <i>Editor</i>	Božidar Potočnik (Univerza v Mariboru, Fakulteta za elektrotehniko, računalništvo in informatiko)
Recenzija <i>Review</i>	Milan Zorman (Univerza v Mariboru, Fakulteta za elektrotehniko, računalništvo in informatiko) Simon Dobrišek (Univerza v Ljubljani, Fakulteta za elektrotehniko)
Lektoriranje <i>Language editing</i>	David Podgorelec (Univerza v Mariboru, Fakulteta za elektrotehniko, računalništvo in informatiko)
Tehnična urednika <i>Technical editors</i>	Jan Perša (Univerza v Mariboru, Univerzitetna založba) Marina Bajić (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>	Virji so lastni, razen če ni navedeno drugače. Avtorji prispevkov in Potočnik (urednik), 2026
Konferenca <i>Conference</i>	Strokovna konferenca ROSUS 2026: Računalniška obdelava slik in njena uporaba v Sloveniji 2026
Datum in kraj konference <i>Conference date & location</i>	19. marec 2026, 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 Leonardis, Simon Dobrišek, Vitomir Štruc, Andrej Šoštarčič, 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, Anže Ferčec, Aleš Holobar, Božidar Potočnik

Založnik **Univerza v Mariboru**
Published by **Univerzitetna založba**
Sloški trg 15, 2000 Maribor, Slovenija
<https://press.um.si>, zalozba@um.si

Izdajatelj **Univerza v Mariboru**
Issued by **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/1022>

Izdano
Published Maribor, Slovenija, marec 2026



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

Besedilo/ *Text* © Avtorji in Potočnik, 2026

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 (20 ; 2026 ; Maribor)

ROSUS 2026 [Elektronski vir] : računalniška obdelava slik in njena uporaba v Sloveniji 2026 : zbornik 20. strokovne konference : [19. marec 2026, Maribor, Slovenija] / urednik Božidar Potočnik. - 1. izd. - E-zbornik. - Maribor : Univerza v Mariboru, Univerzitetna založba, 2026

Način dostopa (URL) : <https://press.um.si/index.php/ump/catalog/book/1022>

ISBN 978-961-299-111-1 (PDF)
doi: 10.18690/um.feri.4.2026
COBISS.SI-ID 270650883

ISBN 978-961-299-111-1 (pdf)
978-961-299-112-8 (mehka vezava)

DOI <https://doi.org/10.18690/um.feri.4.2026>

Cena Brezplačni izvod
Price

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

Citiranje Potočnik, B. (ur.). (2026). *ROSUS 2026 – Računalniška obdelava*
Attribution slik in njena uporaba v Sloveniji 2026: zbornik 20. strokovne konference.
Univerza v Mariboru, Univerzitetna založba. doi:
10.18690/um.feri.4.2026

POKROVITELJI



ROSUS 2026
<http://rosus.feri.um.si>

Kazalo

	Predgovor <i>Foreword</i> Božidar Potočnik	1
VABLJENI PREDAVANJI		3
1	Računalniški vid in umetna inteligenca na slovenski superračunalniški infrastrukturi <i>Computer Vision and AI on Slovenian Supercomputing Infrastructure</i> Janez Perš	5
2	Sistem za samodejno zaznavanje in analizo varnostne razdalje v cestnem prometu <i>System for the Automatic Detection and Analysis of Safe Following Distance in Road Traffic</i> Mitja Pajek	17
DOPOLDANSKA IN POPOLDANSKA SEKCIJA Industrijske aplikacije, medicinske in biomedicinske aplikacije, študentske aplikacije, drugo		23
3	Benchmarking Road Network Extraction Methods for Power Distribution Spatial Planning <i>Primerjava metod ekstrakcije cestnega omrežja za prostorsko načrtovanje distribucije električne energije</i> Janez Križaj, Vitomir Štruc	25
4	Kinematični parametri roke kot kvantitativni označevalci invalidnosti in kognicije pri multipli sklerozi <i>Kinematic Parameters of Hand as Quantitative Markers of Disability and Cognition in Multiple Sclerosis</i> Žan Zajc, Žiga Bizjak, Gregor Brecl Jakob, Miran Buermen, Žiga Špiclin	37

5	Projekt Horizon Europe OnMoveID: Tehnično ozadje avtomatizacije prehoda schengenske meje <i>Horizon Project OnMoveID: Technical Background of Schengen Border Crossing Automation</i> Klemen Vovk, Romanela Lajić, Žan Kogovšek, Jernej Sabadin, Blaž Meden, Žiga Emeršič, Peter Peer	49
6	Rekonstrukcija mišične strukture motoričnih enot iz signalov HD-sEMG z uporabo globokih konvolucijskih nevronske mreže <i>Reconstruction of Muscle Motor Unit Structure From HD-sEMG Signals Using Deep Convolutional Neural Networks</i> Niko Uremović, Niko Lukač, Aleš Holobar	57
7	Decision-Level Fusion of YOLOv8 and PointPillars: Initial Findings <i>Združevanje modelov YOLOv8 in PointPillars na ravni odločanja: Začetne ugotovitve</i> Ivan Vrsalović, Kristijan Lenac	71
8	Razvoj avtonomnega vplivneža z uporabo orodij umetne inteligence <i>Development of an Autonomous Influencer Using Artificial Intelligence Tools</i> Jon Rojnik Goršek, Borut Batagelj	83
9	Personaliziran prikaz digitalnih slik za različne vrste barvne slepote <i>Personalized Digital Image Display for Various Types of Colour Vision Deficiencies</i> Žiga Vinčec, Božidar Potočnik, Martin Šavc	93
10	Human Action Recognition and Custom Dataset for Bus Passenger Safety <i>Prepoznavanje človeških aktivnosti in razvoj podatkovne zbirke za varnost potnikov na avtobusu</i> Dong Gyu Lee, Dong Seog Han	105
11	Zaznava in klasifikacija brezpilotnih letalnikov s kamero in radarjem <i>Detection and Classification of Unmanned Aerial Vehicles With Camera and Radar</i> Pane Andov, Primož Smogavec, Dušan Gleich	115

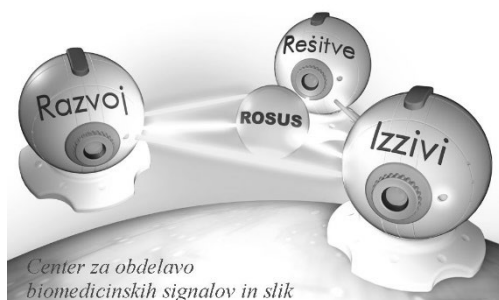
Spoštovani!

BOŽIDAR POTOČNIK
predsednik konference ROSUS 2025

Po devetnajstih konferencah ROSUS 2006–2024 s konferenco ROSUS 2026 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 2026 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 in revijo Ventil, ki so intenzivno promovirali konferenco ROSUS 2026 ter pomen strojnega vida v slovenskem prostoru.



VABLJENI PREDAVANJI

RAČUNALNIŠKI VID IN UMETNA INTELIGENCA NA SLOVENSKI SUPERRAČUNALNIŠKI INFRASTRUKTURI

JANEZ PERŠ

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

Prispevek podaja pregled možnosti za izvajanje računske zahtevnih nalog računalniškega vida in umetne inteligence v Sloveniji na infrastrukturi SLING. Osredotočimo se na dve ključni okolji: petaflopni superračunalnik Vega (240 GPU Nvidia A100 in 960 vozlišč CPU) ter Arnesovo računsko gručo s starejšimi (V100) in novejšimi (H100) GPU za interaktivno in izobraževalno rabo. Opišemo postopke pridobivanja dostopa (testni dostop, razvojni dostop, redni/veliki raziskovalni dostop) in značilne kvote, ki vplivajo na razpoložljivost GPU virov. Pri tem povzamemo tudi vlogo Nacionalnega kompetenčnega centra (EuroCC 2) pri izobraževanjih in podpori uporabnikom. Posebno pozornost namenimo programski združljivosti za računalniški vid (knjižnice PyTorch/TensorFlow, knjižnice CUDA, kontejnerji) ter razlikam med klasičnim paketnim režimom (Slurm) in potrebami sodobnih potekov dela. V zaključku stanje povežemo s projektom Slovenske tovarne umetne inteligence (SLAIF), ki načrtuje nov, za umetno inteligenco optimiziran superračunalnik ter hibridno HPC-oblačno okolje z orkestracijo Kubernetes.

DOI
[https://doi.org/
10.18690/um.feri.4.2026.1](https://doi.org/10.18690/um.feri.4.2026.1)

ISBN
978-961-299-111-1

Ključne besede:
računalniški vid,
umetna inteligenca,
superračunalništvo,
SLING,
Vega,
Arnes,
SLAIF,
Slurm,
Kubernetes

DOI
[https://doi.org/
10.18690/um.feri.4.2026.1](https://doi.org/10.18690/um.feri.4.2026.1)

ISBN
978-961-299-111-1

Keywords:

computer vision,
artificial intelligence,
supercomputing,
SLING,
Vega,
Arnes,
SLAIF,
Slurm,
Kubernetes

COMPUTER VISION AND AI ON SLOVENIAN SUPERCOMPUTING INFRASTRUCTURE

JANEZ PERŠ

University of Ljubljana, Faculty of Electrical Engineering, Ljubljana, Slovenija
janez.pers@fe.uni-lj.si

This paper reviews how computer vision and AI workloads can be executed in Slovenia on the national SLING infrastructure. We focus on two key environments: the EuroHPC petascale system Vega (240 Nvidia A100 GPUs and 960 CPU nodes) and the Arnes HPC cluster, which includes V100 and some H100-based GPU nodes and serves both education and research. We summarise access paths (Test Access, Development Access, Regular/Large Research Access) and the allocation quotas that affect GPU availability for deep learning training. We also note the role of the Slovenian National Competence Centre (EuroCC 2) in training and user support. We discuss software compatibility for CV (PyTorch/TensorFlow, CUDA libraries, containers) and highlight the gap between batch scheduling (Slurm) and modern AI workflows that require interactive development and long-running inference services. Finally, we outline the perspective of the Slovenian AI Factory (SLAIF), which plans an AI-optimised supercomputer and a hybrid HPC-cloud environment with Kubernetes-based orchestration.



University of Maribor Press

1 Uvod

Računalniški vid (RV) je že dolgo eno od osrednjih področij umetne inteligence (UI) – od detekcije objektov in segmentacije do generativnih modelov, ki iz slik, videa ali 3D podatkov izdelujejo nove predstavitve. Skupna značilnost sodobnih pristopov je velika računska zahtevnost: učenje globokih nevronske mreže in transformerjev praviloma zahteva zmogljive grafične pospeševalnike (GPU), dovolj hitro diskovno in omrežno infrastrukturo ter ustrezno programsko okolje.

V Sloveniji je temeljni javni okvir za takšne naloge nacionalna superračunalniška mreža SLING, ki vključuje EuroHPC superračunalnik Vega in Arnesovo računsko gručo, za izbrane projekte pa tudi dostop do dodatnih sistemov prek evropskih mehanizmov odprtega dostopa. Razpoložljivost virov in način pridobivanja dostopa neposredno vplivata na izvedljivost (in časovnico) eksperimentov računalniškega vida, zato sta v prispevku obravnavana skupaj z vprašanjem programske združljivosti.

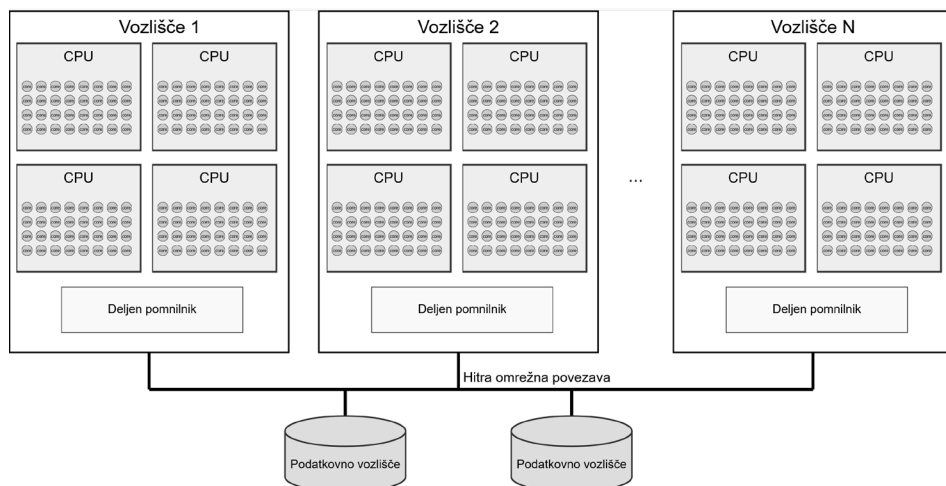
2 Kaj točno je superračunalnik?

Superračunalnik (HPC, angl. high performance computer), prikazan na sliki 1, ni ena “posebna” naprava, temveč velika gruča (angl. cluster) računalniških vozlišč, povezanih z zelo hitrim omrežjem. Vsako vozlišče ima več procesorjev (CPU), pogosto tudi grafične procesne enote (GPU) in deljen pomnilnik. Izjemna zmogljivost ne izhaja iz hitrejših jeder, temveč iz zelo velikega števila jeder, ki lahko računajo vzporedno. Takšna infrastruktura je tehnično zelo podobna tisti pri ponudnikih oblčnih storitev v t.i. podatkovnih centrih (angl. datacenter).

Ključna razlika je v načinu uporabe:

- V oblaku si več uporabnikov hkrati deli dele posameznih vozlišč (virtualizacija).
- Pri superračunalniku pa sistem uporabniku za določen čas dodeli celotna vozlišča ali več vozlišč, medtem ko drugi čakajo v čakalni vrsti.

Zato mora biti problem primeren za masivno vzporedno izvajanje – razdeljen na veliko manjših nalog, ki tečejo hkrati na številnih jedrih ali več GPU.



Slika 1: Arhitektura superračunalnika ni zelo drugačna od arhitekture tipičnega podatkovnega centra, razlika je v načinu dostopa in uporabe.

Vir slike: lasten.

3 Slovenska superračunalniška infrastruktura SLING

Konzorcij SLING (Slovensko nacionalno superračunalniško omrežje) ureja odprti dostop do dela nacionalnih kapacitet in uvaja enoten okvir za prijavo projektov, dodelitev virov in podporo uporabnikom. Ključna prednost takšnega pristopa je, da se lahko uporabnik od začetnega testiranja hitro premakne do večjih dodelitev, hkrati pa se ohranjata sledljivost in transparentnost porabe virov. (SLING, n.d.-b)

3.1 Modeli dostopa in kvote (od testiranja do produkcije)

Razpisi SLING ločijo več tipov dostopa, ki se razlikujejo po namenu in obsegu dodeljenih virov: testni dostop (TA) je namenjen spoznavanju okolja in preverjanju oziroma validaciji programske opreme, razvojni dostop (DA) razvoju in optimizaciji aplikacij, medtem ko sta redni raziskovalni dostop (RRA) in veliki raziskovalni dostop (LRA) predvidena za produkcijsko rabo ter obsežnejše raziskovalne projekte z večjimi dodelitvami virov. Za GPU-intenzivne naloge so pomembne kvote na posamezno dodelitev – npr. v okviru RRA je za Vega-GPU predvidena dodelitev do 25.000 vozliščnih, za Arnes-GPU pa do 10.000 vozliščnih ur na dostop. (SLING, n.d.-b)

V praksi to pomeni, da je smiselno projekt zasnovati iterativno: najprej izvesti manjše teste (npr. preizkus učnih skript, preverjanje odvisnosti, merjenje porabe pomnilnika GPU), nato pa na podlagi rezultatov utemeljiti zahtevo za večjo dodelitev. Tak postopek je pri računalniškem vidu posebej uporaben, ker je strojni učinek (npr. izbira velikosti paketov (batch), uporaba mešane natančnosti, I/O ozka grla pri branju slik ali videa) pogosto odločilen. (SLING, n.d.-b)

3.2 HPC Vega

Vega je petaflopni sistem, ki ga gosti IZUM v Mariboru. Po javno objavljenih specifikacijah ima 960 vozlišč CPU (skupaj 1.920 procesorjev AMD EPYC 7H12 in približno 122.000 jeder) ter 60 vozlišč GPU s skupno 240 grafičnimi pospeševalniki Nvidia A100 s 40 GB VRAM. Zmogljivost je navedena kot 6,9 PFLOPS (trajno) oziroma 10,1 PFLOPS (špice). (IZUM, n.d.-a)

Za računalniški vid je ključna tudi programska oprema: Vega kot delovno okolje navaja podporo za tipične okvirje in knjižnice, potrebne za UI, kot so TensorFlow in PyTorch, ter ekosistem NVIDIA (CUDA, cuDNN, NCCL, TensorRT, DALI, DeepStream). To poenostavi postavitev eksperimentov (zlasti pri uporabi GPU) in zmanjša potrebo po ročnem prevajanju odvisnosti. (IZUM, n.d.-a)

Za reproducibilnost in prenosljivost okolij so pomembni kontejnerji. Dokumentacija sistema navaja, da je na Vegi nameščena in prilagojena platforma Singularity, ki omogoča zaganjanje kontejnerjev v okolju HPC. (HPC Vega documentation, n.d.-b)

Za delitev GPU na manjše enote je pogosto uporabna funkcija Multi-Instance GPU (MIG), zlasti pri manjših modelih RV ali pri inferenci. Dokumentacija Vege navaja, da MIG trenutno ni omogočen, zato je delitev GPU tipično organizirana na ravni dodeljevanja celotnih vozlišč GPU ali GPU-jev znotraj vozlišča. (HPC Vega documentation, n.d.-a)

Uporabniški vmesnik je tradicionalno ukazna vrstica (Slurm), vendar dokumentacija opisuje tudi storitev Open OnDemand, ki omogoča oddajo in spremljanje opravil ter upravljanje datotek prek spletnega brskalnika. (HPC Vega documentation, n.d.-c)

3.3 Arnesova gruča HPC

Arnesova računska gruča je po opisu SLING drugi najzmogljivejši sistem v Sloveniji in je namenjena tako izobraževalnim kot raziskovalnim uporabnikom, pri čemer je med raziskovalnimi področji posebej izpostavljena uporaba v umetni inteligenci in obdelavi velikih podatkov. (SLING, n.d.-a)

Objavljene specifikacije izpostavljajo heterogen nabor vozlišč GPU: 24 vozlišč GPU z 2× Nvidia V100S, 7 vozlišč GPU z 2× Nvidia H100 PCIe in 3 vozlišča GPU s 4× Nvidia H100 SXM5. Sistem uporablja Slurm (navedeni so tudi 'tagi' Slurm za posamezne tipe vozlišč), dostop pa zahteva dvostopenjsko avtentikacijo (2FA). (SLING, n.d.-a)

Tabela 1: Primerjava ključnih okolij za UI/RV v Sloveniji in v projektu SLAIF.

Okolje	Strojna oprema (izbor)	Programski sloj za UI/RV (izbor)	Dostopnost / opombe (izbor)
HPC Vega (IZUM)	60 vozlišč GPU (skupaj 240× Nvidia A100); 960 vozlišč CPU; 6,9 PFLOPS (10,1 PFLOPS peak).	Slurm; TensorFlow, PyTorch; CUDA, cuDNN, NCCL, TensorRT; DALI, DeepStream; Singularity; MIG ni omogočen.	SLING + EuroHPC; RRA do 25.000 GPU-vozliščnih ur; spletni dostop tudi prek Open OnDemand.
HPC Arnes	GPU: 24× (2× V100S), 7× (2× H100 PCIe), 3× (4× H100 SXM5); CPU: heterogen nabor vozlišč EPYC/Xeon.	Slurm; heterogeni profili GPU; primeren za učenje, razvoj in eksperimente; 2FA zahtevan.	razpisi SLING; RRA do 10.000 GPU-vozliščnih ur; pogosto uporabljen v izobraževanju in raziskavah UI.
HPC »Vega 2« (načrtovano v okviru SLAIF)	≈100 PFLOPS (HPC) in ≈10 exaFLOPS (mešana natančnost za UI); ≈2000 pospeševalnikov; 10 PB hitrega diska + 100 PB shrambe Ceph; »cloud« particija.	Hibridni HPC-cloud; poudarek na storitvah in API; orkestracija kontejnerjev s Kubernetes; namenjeno tudi velikim modelom UI in inferenci.	Novi Arnesov podatkovni center pri HE Mariborski otok; ciljna uvedba 2027 (vir navajajo konec leta 2027).

Viri: IZUM (n.d.-a), HPC Vega documentation (n.d.-a; n.d.-b; n.d.-c), SLING (n.d.-a; n.d.-b), IZUM (2025), EuroHPC JU (2025), SLAIF (n.d.), Univerza v Ljubljani (2025).

4 Dostopnost virov in praktični vidik razpoložljivosti

Za uporabnika računalniškega vida je razpoložljivost v praksi kombinacija (i) dodeljene kvote (npr. vozliščne ure na particiji GPU), (ii) politike prioritet in omejitev (časovne omejitve opravil, število vozlišč, število sočasnih opravil), ter (iii)

trenutne zasedenosti sistema. Ker sta Vega in Arnes deljena sistema, je tipičen način dela oddaja opravil v paketni obliki in čakanje v vrsti, kar je bistveno drugače kot pri komercialnih oblakih.

Razpisi SLING so zato zasnovani tako, da podpirajo postopno uvajanje uporabnika: TA in DA sta stalno odprta razpisa z relativno manjšimi dodelitvami, namenjenimi testiranju in razvoju, medtem ko sta RRA in LRA časovno vezana na razpisne roke in omogočata večje dodelitve za produkcijske eksperimente. (SLING, n.d.-b)

Pomemben del dostopnosti je tudi organizacijska in izobraževalna podpora. Slovenski Nacionalni kompetenčni center (NCC) deluje znotraj SLING in je v okviru projekta EuroCC 2 organiziran kot konzorcij partnerjev (Arnes kot vodilni, IJS, univerze in drugi). NCC organizira brezplačna izobraževanja na osnovni, napredni in višji ravni ter pomaga uporabnikom pri uvajanju v okolje HPC. Kot primer specializiranih vsebin sta bila v letu 2025 izvedena »bootcampa« za multi-GPU programiranje in profiliranje programske opreme za UI. (SLING, n.d.-c; EuroCC ACCESS, n.d.; Univerza v Ljubljani FMF, n.d.; IJS Indico, 2025a; IJS Indico, 2025b)

Pri načrtovanju eksperimentov RV je priporočljivo, da se dodelitve vežejo na merljive cilje (npr. število epoh, število poskusov pri optimizaciji hiperparametrov, velikost učne množice) in da se že v fazi TA/DA pripravita avtomatizacija zagona (npr. opravila Slurm) ter zbiranje metrik. S tem se zmanjša tveganje, da bi večja dodelitev RRA/LRA ostala neizkoriščena zaradi programskih ali podatkovnih ozkih grl.

Tipični primeri uporabe superračunalniških virov v računalniškem vidu vključujejo:

- učenje in doučevanje (angl. fine-tuning) globokih modelov na velikih slikovnih/video zbirkah (GPU),
- razširjene eksperimente z več modeli in različnimi hiperparametri (več vzporednih opravil),
- porazdeljeno učenje (multi-GPU) pri večjih modelih ali višjih ločljivostih,
- obsežno inferenco (batch) ali pripravo rezultatov za nadaljnje analize.

5 Programska združljivost za računalniški vid

Združljivost v okolju HPC pomeni predvsem: (i) razpoložljivost ustreznih gonilnikov in okolja CUDA za GPU, (ii) dostop do knjižnic za globoko učenje in računalniški vid, (iii) možnost reprodukcije okolja, ter (iv) učinkovito upravljanje podatkov (lokacija, prepustnost, paralelni dostop).

Vega kot primer navaja podporo za PyTorch in TensorFlow ter nabor knjižnic NVIDIA, ki so pomembne za poteke dela v RV. (IZUM, n.d.-a)

Dokumentacija Vege izrecno opisuje uporabo kontejnerjev Singularity, kar omogoča, da si uporabnik pripravi lastno reproducibilno okolje (npr. natančne verzije PyTorch, CUDA, OpenCV, mmcv itd.) in ga nato zaganja v opravilih Slurm brez poseganja v sistemske module. (HPC Vega documentation, n.d.-b)

Za uporabnike, ki preferirajo grafični dostop, je na Vegi na voljo Open OnDemand, ki omogoča oddajo opravil in upravljanje datotek prek brskalnika. To je uporabno pri učnih delavnicah in pri začetnih fazah projektov, ko je cilj hitra validacija okolja. (HPC Vega documentation, n.d.-c)

6 SLURM kot standard HPC in omejitve pri potekih dela UI

Večina superračunalniških okolij uporablja specializiran razporejevalnik opravil (workload manager), ki skrbi za dodeljevanje virov in čakalne vrste. Slurm je eden najbolj razširjenih: poročila navajajo, da ga uporablja približno polovica lokacij HPC v vzorcu več tisoč sistemov (Hyperion Research, 2023), CNCF pa navaja oceno, da Slurm poganja okoli 65 % sistemov s seznama TOP500 (CNCF, 2024).

Slurm je zasnovan kot upravljalnik virov in razporejevalnik opravil na gručah Linux: dodeli računska vozlišča za določeno časovno obdobje, zažene/spremlja izvajanje in upravlja čakalno vrsto ter politiko deljenja virov. (SchedMD, n.d.)

Takšen model je zelo učinkovit za klasične obremenitve HPC (paketne simulacije, opravila MPI), vendar pri UI pogosto nastopijo dodatne zahteve: interaktivno eksperimentiranje (npr. notebooki), cevovodi, orkestracija več korakov (priprava podatkov – učenje – evalvacija – inferenca) ter zagon dolgoročnih storitev (npr. API

za inferenco). V praksi se zato vse pogosteje kombinira klasični HPC pristop s kontejnerskimi in »cloud-native« pristopi.

Kubernetes je odprtokodni sistem za avtomatizacijo nameščanja, skaliranja in upravljanja »kontejneriziranih« aplikacij in je zaradi svoje zasnove primeren za izvajanje dolgoročnih storitev in platformnih komponent (npr. orodja MLOps). (Kubernetes, n.d.), vendar ga trenutni superračunalniki v Sloveniji ne podpirajo.

7 Perspektiva projekta Slovenske tovarne umetne inteligence (SLAIF)

Marca 2025 je EuroHPC JU objavil izbor dodatnih evropskih tovarn UI, med katerimi je tudi slovenska iniciativa SLAIF. Po objavi bo novi superračunalniški sistem razvijal in upravljal IZUM v sodelovanju z IJS in ARNES, nameščen pa bo v novem podatkovnem centru, ki ga ARNES gradi v bližini hidroelektrarne Mariborski otok. (EuroHPC JU, 2025)

Javni opis SLAIF navaja, da bo novi superračunalnik bistveno zmogljivejši od Vege: okvirno 100 PFLOPS HPC zmogljivosti (okoli $16\times$ Vega) ter za potrebe umetne inteligence okoli 10 exaFLOPS v mešani natančnosti, približno 2000 grafičnih pospeševalnikov, 10 PB hitrega diskovnega sistema za podatkovno intenzivne aplikacije in 100 PB shrambe na tehnologiji Ceph. (IZUM, 2025)

Pomemben konceptualni premik v primerjavi s tradicionalnim HPC je načrtovana delitev na več logičnih particij, vključno s 'cloud services partition', ki je namenjena poganjanju aplikacij UI (inferenca) in drugih dolgoročnih uporabniških storitev. (IZUM, 2025)

V delu, ki se nanaša na sistemsko programsko opremo, je izrecno zapisano, da bodo številni uporabniki potrebovali orkestracijo kontejnerjev, kot je Kubernetes, za skalabilne in upravljive namestitve aplikacij; Kubernetes je tako predviden kot pomemben gradnik pri povezovanju HPC in oblačnih storitev. (IZUM, 2025)

Glede časovnice viri navajajo leto 2027 kot ciljno obdobje prehoda: uradna stran SLAIF navaja, da naj bi bil nov UI-optimiziran sistem v produkciji v drugi polovici leta 2027, medtem ko nekatere partnerske objave omenjajo začetek 2027. V vsakem primeru se Vega predvideva kot ključni premostitveni vir do uvedbe novega sistema. (SLAIF, n.d.; Univerza v Ljubljani, 2025)

8 Zaključek

Slovenska infrastruktura SLING danes omogoča izvajanje nalog računalniškega vida in umetne inteligence na dveh komplementarnih okoljih: Vega nudi petaflopno zmogljivost z A100 GPU in bogatim naborom knjižnic za UI, Arnesova gruča pa predstavlja dostopnejše (tudi izobraževalno) okolje s heterogenimi GPU, vključno z novjšimi H100. Razpoložljivost je določena s kvotami in razpisnimi mehanizmi, zato sta pri večjih eksperimentih ključno načrtovanje in iterativni pristop od TA/DA do RRA/LRA.

SLAIF prinaša jasno usmeritev v UI-optimizirano infrastrukturo, kjer bo poleg klasične HPC zasnove poudarek tudi na storitvah, ki so ključne za UI (dostop prek API, orkestracija kontejnerjev, dolgoročne storitve). Če bodo načrtovane funkcionalnosti (npr. Kubernetes in cloud particija) uresničene, se bo uporabniški model približal sodobnim UI platformam, hkrati pa ohranil prednosti superračunalniškega okolja.

Opomba

Vse navedene tehnične specifikacije in kvote so povzete iz javno objavljenih virov, citiranih v poglavju "Viri in literatura". Ker se sistemi posodablajo, je pri praktični uporabi priporočljivo preveriti zadnjo različico dokumentacije. Prispevek temelji na javno dostopni dokumentaciji in objavah organizacij IZUM, SLING, EuroHPC JU ter na javno dostopnih informacijah projekta SLAIF. Pri virih smo dajali prednost tistim v angleškem jeziku.

Izjava o uporabi orodij umetne inteligence

Ta prispevek je bil pripravljen s pomočjo orodij umetne inteligence, razvitih pri podjetju OpenAI, in sicer ChatGPT (vključno z zmogljivostjo Deep Research ter modelom GPT-5.2 Pro). Za točnost, celovitost in izvirnost prispevka v celoti odgovarja avtor.

Zahvala

Članek je nastal v okviru projekta Slovenska tovarna umetne inteligence (SLAIF – Slovenian AI Factory). Zahvaljujemo se za podporo Evropske komisije / EuroHPC JU ter Ministrstva za visoko šolstvo, znanost in inovacije Republike Slovenije (MVZI) v okviru projekta SLAIF (številka pogodbe o sofinanciranju 101254461).

Literatura

CNCF. (2024, 8 July). Slurm: An HPC workload manager. Cloud Native Computing Foundation. Dostopno na: <https://www.cncf.io/blog/2024/07/08/slurm-an-hpc-workload-manager/> (pridobljeno 14. 2. 2026).

- EuroCC ACCESS. (n.d.). NCC Slovenia. EuroCC ACCESS. Dostopno na: <https://www.eurocc-access.eu/about-us/meet-the-nccs/ncc-slovenia/> (pridobljeno 14. 2. 2026).
- EuroHPC Joint Undertaking. (2025, 12 March). The EuroHPC JU selects additional AI Factories to strengthen Europe's AI leadership. EuroHPC JU. Dostopno na: https://www.eurohpc-ju.europa.eu/eurohpc-ju-selects-additional-ai-factories-strengthen-europes-ai-leadership-2025-03-12_en (pridobljeno 14. 2. 2026).
- HPC Vega documentation. (n.d.-a). Architecture. HPC Vega – IZUM, Maribor, Slovenia. Dostopno na: <https://doc.vega.izum.si/architecture> (pridobljeno 14. 2. 2026).
- HPC Vega documentation. (n.d.-b). Vsebniki (Singularity). HPC Vega – IZUM, Maribor, Slovenia. Dostopno na: <https://si-doc.vega.izum.si/vsebniki/> (pridobljeno 14. 2. 2026).
- HPC Vega documentation. (n.d.-c). Vega Open OnDemand. HPC Vega – IZUM, Maribor, Slovenia. Dostopno na: <https://doc.vega.izum.si/ondemand/> (pridobljeno 14. 2. 2026).
- Hyperion Research. (2023, June). Slurm Remains Top Resource Manager (Special Study Results). Hyperion Research. Dostopno na: <https://hyperionresearch.com/wp-content/uploads/2023/06/Hyperion-Research-MCS-Highlights-Slurm-June-2023.pdf> (pridobljeno 14. 2. 2026).
- IJS Indico. (2025a, 17–18 June). Workshop: EuroCC2 Multi-GPU Programming Bootcamp. IJS Indico. Dostopno na: <https://indico.ijs.si/event/2914/> (pridobljeno 14. 2. 2026).
- IJS Indico. (2025b, 10 July). Bootcamp: Profiling AI Software. IJS Indico. Dostopno na: <https://indico.ijs.si/event/2969/> (pridobljeno 14. 2. 2026).
- Institute of Information Science (IZUM). (2025). Slovenian Artificial Intelligence Factory – SLAIF. IZUM. Dostopno na: https://izum.si/en/ai_factory/ (pridobljeno 14. 2. 2026).
- Institute of Information Science (IZUM). (n.d.-a). Vega. IZUM. Dostopno na: <https://izum.si/en/vega-en/> (pridobljeno 14. 2. 2026).
- Kubernetes. (n.d.). Kubernetes Documentation. Kubernetes. Dostopno na: <https://kubernetes.io/docs/home/> (pridobljeno 14. 2. 2026).
- SchedMD. (n.d.). Overview: Slurm Workload Manager. Slurm Workload Manager Documentation. Dostopno na: <https://slurm.schedmd.com/overview.html> (pridobljeno 14. 2. 2026).
- SLAIF. (n.d.). SLAIF – Slovenian AI Factory (Frequently Asked Questions). SLAIF. Dostopno na: <https://www.slaif.si/> (pridobljeno 14. 2. 2026).
- SLING. (n.d.-a). Arnes HPC cluster. SLING. Dostopno na: <https://www.sling.si/en/arnes-hpc-cluster/> (pridobljeno 14. 2. 2026).
- SLING. (n.d.-b). National calls for open access and use of resources. SLING. Dostopno na: <https://www.sling.si/en/calls-for-open-access-and-use-of-resources/> (pridobljeno 14. 2. 2026).
- SLING. (n.d.-c). National Competence Centre. SLING. Dostopno na: <https://www.sling.si/en/competence-centre/national-competence-centre/> (pridobljeno 14. 2. 2026).
- Univerza v Ljubljani. (2025, 14 March). The University of Ljubljana will play a key role in establishing the Slovenian Artificial Intelligence Factory. Univerza v Ljubljani. Dostopno na: <https://www.uni-lj.si/en/news/2025-03-14-the-university-of-ljubljana-will-play-a-key-role-in-establishing-the-slovenian-artificial-intelligence-factory> (pridobljeno 14. 2. 2026).
- Univerza v Ljubljani FMF. (n.d.). EUROCC2 - National Competence Centres in the framework of EuroHPC. University of Ljubljana, Faculty of Mathematics and Physics. Dostopno na: <https://www.fmf.uni-lj.si/sl/raziskave/mednarodni/eurocc2/> (pridobljeno 14. 2. 2026).

SISTEM ZA SAMODEJNO ZAZNAVANJE IN ANALIZO VARNOSTNE RAZDALJE V CESTNEM PROMETU

MITJA PAJEK

Tri Sense, Ljubljana, Slovenija
mitja.pajek@tri-sense.com

Tri Sense d. o. o. je slovensko tehnološko podjetje, specializirano za strojni vid in rešitve na osnovi umetne inteligence za zahtevna realna okolja. Na konferenci bo predstavljen sistem za samodejno zaznavanje in analizo varnostne razdalje v cestnem prometu, ki temelji na kombinaciji kamer, naprednih algoritmov računalniškega vida in umetne inteligence. Rešitev omogoča natančno zaznavanje vozil, merjenje medsebojnih razdalj ter identifikacijo nevarnih situacij v realnem času brez poseganja v obstoječo cestno infrastrukturo. Sistem je zasnovan za dolgoročno, neprekinjeno delovanje v zunanjih okoljih in omogoča analizo prometne varnosti ter podporo preventivnim ukrepom in odločanju upravljavcev cest. Rešitev prispeva k zmanjševanju prometnih nesreč ter predstavlja praktičen primer uporabe umetne inteligence za izboljšanje varnosti in učinkovitosti prometnih sistemov.

DOI

[https://doi.org/
10.18690/um.feri.4.2026.2](https://doi.org/10.18690/um.feri.4.2026.2)

ISBN

978-961-299-111-1

Ključne besede:

računalniški vid v prometu,
realnočasovno zaznavanje
vozil,
merjenje medvozične
razdalje,
inteligentni prometni
sistemi,
avtomatiziran nadzor
prometne varnosti



Univerzitetna založba
Univerze v Mariboru

DOI
[https://doi.org/
10.18690/um.feri.4.2026.2](https://doi.org/10.18690/um.feri.4.2026.2)

ISBN
978-961-299-111-1

Keywords:

computer vision in traffic,
real-time vehicle detection,
inter-vehicle distance
measurement,
intelligent transportation
systems,
automated traffic safety
monitoring

SYSTEM FOR THE AUTOMATIC DETECTION AND ANALYSIS OF SAFE FOLLOWING DISTANCE IN ROAD TRAFFIC

MITJA PAJEK

Tri Sense, Ljubljana, Slovenia
mitja.pajek@tri-sense.com

Tri Sense d.o.o. is a Slovenian technology company that specialises in providing machine vision and artificial intelligence-based solutions for challenging real-world environments. At the conference, we will present a system for automatically detecting and analysing safe following distances in road traffic. This system is based on a combination of cameras, advanced computer vision algorithms, and artificial intelligence. The solution can accurately detect vehicles, measure inter-vehicle distances, and identify hazardous situations in real time without interfering with existing road infrastructure. Designed for long-term, continuous outdoor operation, the system enables traffic safety analysis, supports preventive measures, and informs decision-making by road authorities. This solution helps to reduce traffic accidents and demonstrates how artificial intelligence can be used to improve the safety and efficiency of transport systems.



University of Maribor Press

1 Uvod

V prispevku je predstavljen sistem za samodejno zaznavanje in analizo varnostne razdalje v cestnem prometu, ki ga je razvilo podjetje Tri Sense d. o. o. Sistem temelji na lastnem razvoju in domačem znanju ter je zaščiteno s patentno prijavo. Njegov primarni namen je povečanje prometne varnosti z aktivnim vplivanjem na vedenje voznikov in spodbujanjem vzdrževanja ustrezne varnostne razdalje med vozili.

Predstavljena rešitev uvaja inteligentni prometni sistem, ki za zaznavanje in analizo prometnih tokov uporablja metode računalniškega vida. Sistem temelji na kombinaciji kamer in naprednih algoritmov za obdelavo slik, detekcijo objektov ter ocenjevanje razdalj in hitrosti vozil. Algoritmi v realnem času zaznavajo vozila, določajo njihovo medsebojno razdaljo ter na podlagi izračunanih parametrov ocenjujejo ustreznost varnostne razdalje glede na prometno situacijo. Slika 1 prikazuje primer namestitve inteligentnega prometnega sistema na obstoječo ulično svetilko, kar omogoča neinvazivno integracijo v prometno infrastrukturo brez dodatnih posegov v vozišče.

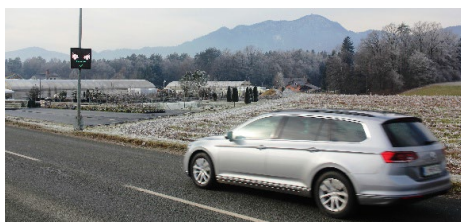


Slika 1: Sistem za samodejno zaznavanje in analizo varnostne razdalje, nameščen na ulično svetilko: kamera (zgoraj) in omarica z računalnikom ter hranilnikom energije (spodaj).

Vir: lasten.

Eden ključnih ciljev sistema je zmanjšati število nevarnih voženj preblizu za drugim vozilom (t. i. "tailgating"), ki predstavlja enega pomembnejših dejavnikov tveganja za nastanek naletnih prometnih nesreč. Z zmanjšanjem tovrstnega vedenja se posledično zmanjšuje število prometnih nesreč, hkrati pa se izboljšujeta pretočnost in stabilnost prometnega toka, kar lahko prispeva tudi k zmanjšanju zastojev.

Načelo delovanja sistema temelji na neprekinjenem spremljanju prometnega toka. Sistem kontinuirano meri medvozilne razdalje in ocenjuje hitrost gibanja vozil, nato pa voznikom v realnem času posreduje povratno informacijo o ustreznosti varnostne razdalje preko LED-prikazovalnika ob cesti. Prikazovalnik z jasno vizualno signalizacijo opozarja na neustrezno razdaljo oziroma potrjuje ustrezno varnostno razdaljo. Slika 2 prikazuje primer izpisa na LED-prikazovalniku v različnih prometnih situacijah; v prvi vrstici sta prikazana primera neustrezne, v drugi vrstici pa primera ustrezne varnostne razdalje med vozili.



Slika 2: Prikazovalnik LED, namenjen prikazovanju ustreznosti varnostne razdalje: neustrezna (vrstica 1) in ustrezna varnostna razdalja med vozili (vrstica 2).

Vir: lasten.

Sistem je zasnovan za dolgoročno in zanesljivo delovanje v zunanjih okoljih ter je primeren za namestitev na lokacijah z visoko prometno obremenitvijo ali na odsekih kritične infrastrukture, kjer je tveganje za prometne nesreče in zastoje povečano. Poleg neposrednega opozarjanja voznikov omogoča tudi zajem, shranjevanje in nadaljnjo analizo prometnih podatkov. Zbrani podatki se lahko uporabijo za statistične obdelave, analizo prometne varnosti ter podporo odločanju upravljavcev cest pri načrtovanju preventivnih in infrastrukturnih ukrepov.

Predstavljena rešitev tako predstavlja praktičen primer implementacije umetne inteligence v realnem prometnem okolju, kjer s kombinacijo zaznavanja, analitike in vizualnega obveščanja neposredno vpliva na vedenje voznikov ter prispeva k večji varnosti in učinkovitosti prometnih sistemov.

DOPOLDANSKA IN POPOLDANSKA SEKCIJA

Industrijske aplikacije
Medicinske in biomedicinske aplikacije
Študentske aplikacije
Drugo

BENCHMARKING ROAD NETWORK EXTRACTION METHODS FOR POWER DISTRIBUTION SPATIAL PLANNING

JANEZ KRIŽAJ, VITOMIR ŠTRUC

University of Ljubljana, Faculty of Electrical Engineering, Slovenia
janez.krizaj@fe.uni-lj.si, vitomir.struc@fe.uni-lj.si

Accurate road network extraction from high-resolution overhead imagery is a prerequisite for corridor-aware routing and cost modelling in power-system infrastructure planning, where medium- and low-voltage lines often follow transport rights-of-way. Yet roads are thin, occluded, and cluttered, and downstream optimisation needs routable graphs, not just masks. We benchmark six representative pipelines (SAM-Road, D-LinkNet, CRESI, CU-dGCN, Sat2Graph, and U-Net-ResNet18) on SpaceNet imagery. The comparison reveals trade-offs between segmentation-centric and graph-aware designs and highlights where topology fails most often. Finally, practical guidance is provided for selecting road extractors that deliver reliable, optimisation-ready networks for power distribution corridor planning and the routing of medium- and low-voltage lines.

DOI

[https://doi.org/
10.18690/um.feri.4.2026.3](https://doi.org/10.18690/um.feri.4.2026.3)

ISBN

978-961-299-111-1

Keywords:

road network extraction,
road graph extraction,
satellite imagery,
benchmarking,
network connectivity,
power distribution spatial
planning



University of Maribor Press

DOI
[https://doi.org/
10.18690/um.feri.4.2026.3](https://doi.org/10.18690/um.feri.4.2026.3)

ISBN
978-961-299-111-1

Ključne besede:
ekstrakcija cestnega
omrežja,
satelitske slike,
primerjalno ovrednotenje,
analiza povezljivosti,
prostorsko načrtovanje
distribucijskega
elektroenergetskega omrežja

PRIMERJAVA METOD EKSTRAKCIJE CESTNEGA OMREŽJA ZA PROSTORSKO NAČRTOVANJE DISTRIBUCIJE ELEKTRIČNE ENERGIJE

JANEZ KRIŽAJ, VITOMIR ŠTRUC

Univerza v Ljubljani, Fakulteta za elektrotehniko, Slovenija
janez.krizaj@fe.uni-lj.si, vitimir.struc@fe.uni-lj.si

Natančna ekstrakcija cestnega omrežja iz satelitskih posnetkov omogoča določanje poteka in modeliranje stroškov pri načrtovanju infrastrukture elektroenergetskega sistema, kjer srednje- in nizkonapetostni koridorji pogosto sledijo cestam. Ker so ceste na satelitskih slikah pogosto ozke, deloma prikrite in nejasne, lahko manjše nepravilnosti v segmentacijskih maskah po vektorski pretvorbi povzročijo prekinjeno povezljivost, zato je za načrtovanje bolj koristno uporabiti grafe, ki omogočajo načrtovanje poti. Na zbirki SpaceNet smo primerjali šest reprezentativnih postopkov (SAM-Road, D-LinkNet, CRESI, CU-dGCN, Sat2Graph in U-Net-ResNet18). V primerjavi proučimo razlike med klasičnimi segmentacijskimi postopki in novejšimi postopki z grafnimi zasnovami ter analiziramo, kje posamezni postopki najpogosteje odpovejo. Navedemo tudi praktična navodila za izbiro ustreznih postopkov, ki zagotavljajo zanesljivo detekcijo cest, primerno za optimizacijo pri načrtovanju srednje- in nizko-napetostnih vodov v distribucijskem električnem omrežju.



Univerzitetna založba
Univerze v Mariboru

1 Introduction

Accurate road network extraction from aerial and satellite imagery is a longstanding challenge in computer vision and remote sensing. It underpins crucial applications in navigation, autonomous driving, urban planning, and disaster response, and it is also a key input for the spatial planning of power system infrastructure. In distribution network development, candidate medium- and low-voltage corridors are often derived from transport right-of-way: routing along roads can reduce permitting effort, limit excavation on private parcels, and guarantee construction and maintenance access. Consequently, the quality and completeness of the road layer can directly affect corridor-aware routing, cost modelling, and feasibility screening of candidate alignments.

While early approaches relied on manual mapping or classical image processing, the advent of deep learning has led to significant progress in automated road extraction. Yet, producing routable road networks – continuous centerline graphs suitable for path planning – remains difficult. CNN-based segmentation often yields fragmented or noisy road masks, necessitating post-processing (e.g. thinning or connectivity analysis) to derive vector road graphs. Importantly, errors that are visually minor at the pixel level (small gaps, slightly misplaced intersections) can become major issues once converted into graphs, by disconnecting otherwise feasible corridors or introducing spurious shortcuts.

Motivated by these challenges, this paper conducts a comparative evaluation of modern road extraction techniques, covering both pixel-level and topology-aware paradigms. We focus on state-of-the-art models that span the spectrum from pure mask prediction to approaches that explicitly model connectivity, with the goal of clarifying how structural priors affect accuracy and practical suitability for optimisation-ready workflows in infrastructure planning.

The contributions of this work are as follows:

- Clear taxonomy of road extraction methods: We outline a taxonomy dividing approaches into segmentation-based, skeleton-based, and graph-based methods, informed by recent surveys (Chen et al., 2022) (Sharma et al., 2024) (Liu et al., 2024) (Wang et al., 2025). This categorization highlights the field's

evolution from pixel-level segmentation towards representations that capture road connectivity and topology.

- Systematic evaluation of representative models: We implement and evaluate six representative pipelines: D-LinkNet (Zhou et al., 2018) and U-Net–ResNet18 (Robinson et al., 2024) as pure segmentation baselines, CU-dGCN (Vekinis, 2023) as a segmentation model augmented with graph reasoning, CRESI (Etten, 2020) as a segmentation+skeleton pipeline, Sat2Graph (He et al., 2020) as a direct graph prediction approach, and SAM-Road (Hetang et al., 2024) as a foundation-model-based method that combines mask prediction with a lightweight graph module.
- Insights into topology-aware extraction: Through qualitative and quantitative analysis, we examine the ability of each method to produce coherent road networks. We show how methods that leverage graph structure (either via architectural components or post-processing) improve topological correctness (fewer broken segments and spuriously disconnected roads) compared to vanilla segmentation.
- Practical considerations for spatial optimisation: We relate the reported road-network metrics to requirements that arise in corridor-aware planning (e.g., avoiding disconnected corridors and unrealistic shortcuts), with emphasis on the needs of power system infrastructure planning.

Overall, our work demonstrates how increasingly sophisticated models address the limitations of pure segmentation, and it offers guidance for selecting road extraction pipelines when the end goal is a connected, routable network for downstream planning workflows. The next sections present related work, describe the evaluated methods, and report experimental results.

2 Related Work

Road extraction from remote sensing imagery has been widely studied, and recent surveys document a clear shift from classical image processing toward deep learning pipelines (Chen et al., 2022) (Sharma et al., 2024) (Liu et al., 2024) (Wang et al., 2025). Consistent with these surveys, we adopt a taxonomy that groups approaches into: (i) segmentation-based methods that predict pixel-wise road masks, (ii) skeleton-based pipelines that derive centerlines/graphs from predicted masks via thinning and pruning, and (iii) graph-based approaches that model connectivity explicitly, either

by learning on graphs or by directly predicting graph structures. This taxonomy is practically relevant whenever the downstream goal is a routable network: small gaps or spurious links that are visually minor at the pixel level can disconnect feasible routes or introduce unrealistic shortcuts once a mask is converted into a graph.

Segmentation-based approaches commonly cast road extraction as semantic segmentation with encoder–decoder CNNs (e.g., FCN/U-Net/SegNet families) (Badrinarayanan et al., 2017) (Chen et al., 2018), producing a dense road mask that can be thresholded and later vectorized. D-LinkNet (Zhou et al., 2018) improves context aggregation (e.g., via dilated convolutions), while lightweight variants such as U-Net–ResNet18 (Robinson et al., 2024) remain strong baselines. A key limitation for optimisation-ready use is that visually minor mask defects (gaps, junction artifacts) can become major topological errors after skeletonization.

Skeleton-based pipelines address this limitation by augmenting segmentation with explicit mask cleanup, skeletonization, and pruning to improve routability; a representative example is CRESI (Etten, 2020). While effective in practice, their final graph quality remains sensitive to the continuity of the predicted mask and to heuristic post-processing choices.

Graph-based methods target connectivity more directly. Some incorporate topology cues into learning, for example via multi-task designs that add structural supervision such as orientations (Batra et al., 2019), while hybrids inject graph reasoning modules to propagate connectivity information (Vekinis, 2023). In parallel, graph-centric approaches avoid heuristic vectorization by predicting road graphs explicitly: RoadTracer (Bastani et al., 2018) pioneered iterative graph construction, Sat2Graph (He et al., 2020) predicts vertices and edges in a single forward pass, and transformer-based detectors such as RNGDet (Xu et al., 2022) and RNGDet++ (Xu et al., 2023) extend this paradigm with detection-style graph decoding. This trend toward graph outputs (rather than raster masks) is also emphasized in recent surveys (Chen et al., 2022) (Liu et al., 2024) (Wang et al., 2025).

3 Methods

We compare six representative approaches spanning the three categories introduced above: segmentation-based mask predictors (D-LinkNet, UNet-ResNet18), a skeleton-based pipeline that derives a centerline graph from a predicted mask

(CRESI), and graph-based methods that explicitly model connectivity - either by refining mask predictions with learned graph reasoning (CU-dGCN, SAM-Road) or by directly predicting a road graph (Sat2Graph). Below we summarize each model's architecture and output format.

3.1 U-Net–ResNet18 (Segmentation baseline)

U-Net–ResNet18 (Robinson et al., 2024) employs a standard U-Net encoder-decoder architecture with a ResNet-18 backbone, providing a lightweight yet effective road segmentation baseline. As a pure segmentation approach, it outputs a binary road mask after thresholding; connectivity is only implicitly learned and not explicitly enforced.

3.2 D-LinkNet (Segmentation baseline)

D-LinkNet (Zhou et al., 2018) is a CNN road segmenter based on an encoder-decoder architecture with skip connections. Dilated convolutions enlarge the receptive field while preserving spatial detail, which helps bridge gaps caused by occlusions or low contrast. In typical configurations, a pretrained ResNet backbone extracts multi-scale features and the decoder fuses them into a dense road-probability map. The output is a binary mask after thresholding; connectivity is encouraged only implicitly through context aggregation and is not explicitly enforced.

3.3 CRESI (Segmentation + skeleton)

CRESI (Etten, 2020) targets large-scale road graph extraction through a pragmatic multi-stage pipeline. A segmentation model predicts road likelihood on overlapping tiles, which are then mosaicked into a global mask to reduce boundary artifacts. Post-processing cleans the mask and fills small gaps before skeletonization. The skeleton is converted into a graph, and pruning removes spurs and short branches to improve routability. This design emphasizes that the graph-construction stage can be as important as the segmentation backbone.

3.4 CU-dGCN (Segmentation + graph reasoning)

CU-dGCN (Vekinis, 2023) couples convolutional image features with graph neural network reasoning. A road centerline graph is constructed or predicted and then refined via message passing, allowing connectivity cues to propagate across gaps and complex junctions. The refined graph representation is mapped back to improve the road prediction, effectively combining local appearance with global structure. This hybrid design aims to deliver better routability without fully abandoning the efficiency and stability of mask-based segmentation.

3.5 SAM-Road (Segmentation + graph reasoning)

SAM-Road (Hetang et al., 2024) adapts the Segment Anything Model (SAM) to road extraction using road-specific prompting strategies and a decoding pipeline that yields road masks and graph representations. Leveraging a foundation-model prior can help under domain shift and with limited labeled data. In practice, SAM-Road still depends on prompt design and decoding choices to avoid fragmented masks and to produce a routable network, but it reduces reliance on training a task-specific segmenter from scratch.

3.6 Sat2Graph (Direct graph prediction)

Sat2Graph (He et al., 2020) predicts a road graph in a single forward pass by learning structured representations of vertices and edges from the image. Compared to tracing, global prediction is less sensitive to local dead-ends and can better preserve long-range continuity. The approach relies on carefully defined training targets and a robust graph decoding stage to avoid missing thin roads or introducing shortcuts. The output is a graph, which can be rasterized to a mask for overlap-based evaluation when needed.

4 Experiments

We compare representative segmentation-first and graph-centric road extraction methods on the SpaceNet roads benchmark, which provides satellite imagery paired with ground-truth vector road graphs. Because our target downstream use is corridor-aware routing for spatial optimisation (including power system

infrastructure planning), we focus on connectivity and routability rather than pixel overlap alone. To make the comparison consistent across output types, mask predictions are thresholded and converted into centerline graphs using a standard thinning-and-tracing pipeline with light cleanup, while graph predictions are evaluated directly after minor snapping of near-coincident vertices. We report TOPO¹ precision/recall/F1 and APLS (Average Path Length Similarity) following the SpaceNet evaluation definitions; in Table 1 the precision/recall/F1 values correspond to TOPO rather than pixel-level segmentation scores.

Table 1: SpaceNet results for representative methods. Prec./Rec./F1 are TOPO metrics; APLS measures routing fidelity. Values are in %.

Method	Prec. ↑	Rec. ↑	F1 ↑	APLS ↑
U-Net–ResNet18	68.96	66.32	67.61	53.77
D-LinkNet	82.42	60.06	68.80	56.96
CRESI	79.11	69.47	71.87	56.38
CU-dGCN	62.13	61.66	62.32	58.19
Sat2Graph	85.93	76.55	80.97	64.43
SAM-Road	93.03	70.97	80.52	71.64

Table 1 summarizes quantitative results. For corridor selection and cost modelling, APLS is particularly informative because it measures shortest-path similarity and is therefore sensitive to disconnections that would force unrealistic detours in an optimised route. TOPO complements APLS by capturing local structural agreement around intersections and branches: high precision discourages spurious links (which can create infeasible shortcuts along non-existent roads), while high recall reduces the risk of missing feasible corridors. Among the compared methods, SAM-Road achieves the highest APLS and very high TOPO precision, indicating clean graph predictions with comparatively few false connections, but its lower recall suggests that it can miss parts of the network. Sat2Graph provides the best balance in TOPO and also strong APLS, which can be advantageous when the planning goal is to enumerate alternative access corridors rather than only the most certain ones. CU-dGCN and CRESI sit in the mid-range, while the segmentation-first baselines (D-LinkNet and U-Net–ResNet18) exhibit lower APLS, consistent with the fact that small mask gaps and junction errors tend to be amplified during vectorization. Overall, the table suggests that graph-centric approaches are better suited for automation-heavy planning workflows, whereas pure segmentation outputs are

¹ TOPO evaluates the similarity of reachable subgraphs from randomly sampled vertices in the predicted road graph and ground truth.

more likely to require additional post-processing or manual GIS corrections before they can reliably support corridor-aware routing.

Fig. 1 provides a qualitative comparison on a challenging orthophoto from Surveying and Mapping Authority of the republic of Slovenia (hereinafter GURS), where we use the OSM road layer (OpenStreetMap contributors, 2017) as a pseudo ground truth reference for visual assessment. Under this interpretation, segments present in OSM but missing in the prediction correspond to false negatives (potentially removing feasible corridors), whereas predicted segments absent in OSM correspond to false positives (potentially creating spurious, non-existent shortcuts). We note that OSM itself can be incomplete or slightly misaligned, so some apparent false positives may in reality be unmapped roads; nevertheless, for planning workflows that start from an OSM-like baseline, such deviations would still require additional verification.

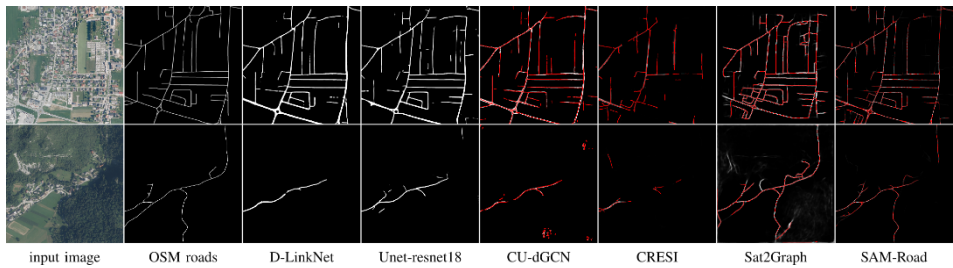


Figure 1: Visual comparison of road extraction outputs. Column 1: GURS orthophoto (GURS, 2026). Column 2: OSM road layer rasterized as a binary mask. Remaining columns: method-predicted road masks; where a road graph/centerline is natively available, it is drawn as red lines over the mask.

Source: GURS and author-generated images.

In the urban part (top of Fig. 1), D-LinkNet largely follows the main OSM corridors but produces widened road regions and small spurious blobs that would likely turn into extra branches after thinning. The U-Net baseline is noisier, deviating from the OSM topology with more fragmented and spurious segments. CU-dGCN and CRESI suppress many of these artifacts and better preserve the mapped connectivity, but they still exhibit breaks where OSM indicates continuous roads, especially near shadows and vegetation boundaries.

In the rural/forested part (bottom of Fig. 1), the OSM road is partially obscured by canopy and overgrowth in the orthophoto. Here, Sat2Graph appears to recover a more continuous connection along the OSM corridor than the other methods, suggesting improved robustness to partial occlusion in the tested GURS imagery. A plausible explanation is that Sat2Graph's direct graph supervision and global decoding impose a stronger connectivity prior and may generalize better under dataset shift between its training data and national orthophotos. SAM-Road yields a clean, high-precision network that aligns well with the main OSM roads, but it can still break under heavy occlusion and may miss minor branches, which is consistent with the precision--recall trade-off observed in Table 1.

5 Conclusion

This paper organized road extraction methods for satellite imagery by output representation and compared representative approaches using graph-oriented evaluation. The main takeaway is that road masks that look plausible are not necessarily routable once converted to a centerline graph: small gaps and junction errors can dominate final network quality and are better captured by TOPO and APLS than by pixel overlap alone. This distinction matters directly for power system spatial planning, where road networks are commonly used as candidate corridors and access constraints for routing underground cables or overhead lines. In this setting, disconnected graphs can eliminate feasible corridors, while spurious links can lead to unrealistic low-cost routes that would later fail during permitting or field verification.

Across the compared methods, the quantitative results indicate that graph-centric approaches (SAM-Road and Sat2Graph) provide higher routing fidelity (APLS) and more stable connectivity than segmentation-first baselines, making them more suitable for automation-heavy corridor-aware optimisation. The qualitative comparison on GURS orthophotos further highlights that OSM roads provide a valuable prior but can be incomplete in rural or rapidly changing areas, and vegetation occlusion remains a key failure mode. In our tested examples, Sat2Graph appears more robust in recovering roads that are partially hidden by overgrowth, which may stem from its direct graph supervision and stronger generalization under domain shift.

Acknowledgement

The authors gratefully acknowledge financial support from the fundamental research project Opti-AI (Optimal Integration of Power System Infrastructure through AI-based Spatial Planning, project no. J2-60030), funded by the Slovenian Research and Innovation Agency (ARIS).

Literatura

- Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 39, 2481–2495.
- Bastani, F., He, S., Abbar, S., Alizadeh, M., Balakrishnan, H., Chawla, S., Madden, S., & DeWitt, D. (2018). Roadtracer: Automatic extraction of road networks from aerial images. *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, (str. 4720–4728).
- Batra, A., Singh, S., Pang, G., Basu, S., Jawahar, C. V., & Paluri, M. (2019). Improved road connectivity by joint learning of orientation and segmentation. *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, (str. 10385–10393).
- Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. *Proc. Eur. Conf. Comput. Vis. (ECCV)*, (str. 801–818).
- Chen, Z., Deng, L., Luo, Y., Li, D., Junior, J. M., Gonçalves, W. N., Nurunnabi, A. A., Li, J., Wang, C., & Li, D. (2022). Road extraction in remote sensing data: A survey. *Int. J. Appl. Earth Obs. Geoinf.*, 112, 102833. <https://doi.org/https://doi.org/10.1016/j.jag.2022.102833>
- Etten, A. V. (2020). City-Scale Road Extraction from Satellite Imagery v2: Road Speeds and Travel Times. *2020 IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, (str. 1775–1784). <https://doi.org/10.1109/WACV45572.2020.9093593>
- GURS. (2026). National orthophoto dataset. *Surveying and Mapping Authority of the Republic of Slovenia*. Ministry of Natural Resources and Spatial Planning. <https://podatki.gov.si/dataset/ortofoto>
- He, S., Bastani, F., Jagwani, S., Alizadeh, M., Balakrishnan, H., Chawla, S., Elsharif, M. M., Madden, S., & Sadeghi, M. A. (2020). Sat2Graph: Road Graph Extraction through Graph-Tensor Encoding. *Proc. Eur. Conf. Comput. Vis. (ECCV)*. <https://api.semanticscholar.org/CorpusID:220647459>
- Hetang, C., Xue, H., Le, C., Yue, T., Wang, W., & He, Y. (2024). Segment anything model for road network graph extraction. *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, (str. 2556–2566).
- Liu, R., Wu, J., Lu, W., Miao, Q., Zhang, H., Liu, X., Lu, Z., & Li, L. (2024). A Review of Deep Learning-Based Methods for Road Extraction from High-Resolution Remote Sensing Images. *Remote Sens.*, 16. <https://doi.org/10.3390/rs16122056>
- OpenStreetMap contributors. (2017). Planet dump retrieved from <https://planet.osm.org>. *Planet dump retrieved from https://planet.osm.org*.
- Robinson, C., Corley, I., Ortiz, A., Dodhia, R., Ferres, J. M., & Najafirad, P. (2024). Seeing the roads through the trees: A benchmark for modeling spatial dependencies with aerial imagery. *Seeing the roads through the trees: A benchmark for modeling spatial dependencies with aerial imagery*. <https://arxiv.org/abs/2401.06762>
- Sharma, P., Kumar, R., Gupta, M., & Nayyar, A. (2024). A critical analysis of road network extraction using remote sensing images with deep learning. *Spat. Inf. Res.*, 32, 485–495. <https://doi.org/10.1007/s41324-024-00576-y>
- Vekinis, A. A. (2023). Graph Reasoned Multi-Scale Road Segmentation in Remote Sensing Imagery. *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, (str. 6890–6893). <https://doi.org/10.1109/IGARSS52108.2023.10281660>
- Wang, X., Jin, X., Dai, Z., Wu, Y., & Chehri, A. (2025). Deep Learning-Based Methods for Road Extraction From Remote Sensing Images: A vision, survey, and future directions. *IEEE Geosci. Remote Sens. Mag.*, 13, 55–78. <https://doi.org/10.1109/MGRS.2024.3491014>

- Xu, Z., Liu, Y., Gan, L., Sun, Y., Wu, X., Liu, M., & Wang, L. (2022). Rngdet: Road network graph detection by transformer in aerial images. *IEEE Trans. Geosci. Remote Sens.*, 60, 1–12.
- Xu, Z., Liu, Y., Sun, Y., Liu, M., & Wang, L. (2023). Rngdet++: Road network graph detection by transformer with instance segmentation and multi-scale features enhancement. *IEEE Robot. Autom. Lett.*, 8, 2991–2998.
- Zhou, L., Zhang, C., & Wu, M. (2018). D-LinkNet: LinkNet with Pretrained Encoder and Dilated Convolution for High Resolution Satellite Imagery Road Extraction. *2018 IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, (str. 192-1924).
<https://doi.org/10.1109/CVPRW.2018.00034>

KINEMATIČNI PARAMETRI ROKE KOT KVANTITATIVNI OZNAČEVALCI INVALIDNOSTI IN KOGNICIJE PRI MULTIPLI SKLEROZI

ŽAN ZAJC,¹ ŽIGA BIZJAK,¹ GREGOR BRECLJAKOB,²
MIRAN BUERMEN,¹ ŽIGA ŠPICLIN¹

¹ Univerza v Ljubljani, Fakulteta za elektrotehniko, Ljubljana, Slovenija
zz88301@student.uni-lj.si, ziga.bizjak@fe.uni-lj.si, miran.burmen@fe.uni-lj.si,
ziga.spiclin@fe.uni-lj.si

² Univerzitetni klinični center Ljubljana, Nevrološka klinika, Ljubljana, Slovenija
gregor.brecljakob@kclj.si

Nevrološke bolezni, kot je multipla skleroza (MS), pogosto povzročajo motnje fine motorike, zato so potrebne objektivne metode za njihovo ocenjevanje. V tej študiji smo razvili stereo-vizualni sistem za sledenje dlani, ki omogoča 3D rekonstrukcijo trajektorij 21 anatomskih točk roke med izvajanjem testa devetih zatičev (NHPT). Na podlagi rekonstruiranih gibanj smo izračunali kinematične parametre in jih klinično ovrednotili na vzorcu 162 bolnikov z MS z analizo povezav z razširjeno lestvico prizadetosti (EDSS) in kognitivno oceno CogEval. Rezultati so pokazali, da je višja stopnja nevrološke prizadetosti najmočnejše povezana s podaljšanim časom izvedbe NHPT ter s spremenjenimi gibalnimi vzorci palca, kot sta večja linearnost in manjša sunkovitost gibanja. Pri kognitivnem statusu se je kot ključni dejavnik izkazala hitrost izvedbe naloge. Predlagani pristop je neinvaziven in cenovno dostopen ter omogoča poglobljen vpogled v kakovost gibanja, kar predstavlja obetavno orodje za objektivno in skalabilno spremljanje motoričnih sprememb pri MS.

DOI

[https://doi.org/
10.18690/um.feri.4.2026.4](https://doi.org/10.18690/um.feri.4.2026.4)

ISBN

978-961-299-111-1

Ključne besede:

stereo video-analiza,
kinematika roke,
sledenje anatomskim
točkam,
ocena fine motorične
sposobnosti,
multipla skleroza,
klinična ocena,
digitalni biomarkerji



Univerzitetna založba
Univerze v Mariboru

Keywords:

stereo video analysis,
hand kinematics,
anatomical landmark
tracking,
fine motor skill assessment,
multiple sclerosis,
clinical assessment,
digital biomarkers

KINEMATIC PARAMETERS OF HAND AS QUANTITATIVE MARKERS OF DISABILITY AND COGNITION IN MULTIPLE SCLEROSIS

ŽAN ZAJC,¹ ŽIGA BIZJAK,¹ GREGOR BRECLJAKOB,²
MIRAN BUERMEN,¹ ŽIGA ŠPICLIN¹

¹ University of Ljubljana, Faculty of Electrical Engineering, Ljubljana, Slovenia
zz88301@student.uni-lj.si, ziga.bizjak@fe.uni-lj.si, miran.burmen@fe.uni-lj.si,
ziga.spiclin@fe.uni-lj.si

² University Medical Center Ljubljana, Division of Neurology, Ljubljana, Slovenia
gregor.brecljakob@kclj.si

Neurological diseases such as multiple sclerosis (MS) frequently lead to impairments in fine motor function, highlighting the need for objective assessment methods. In this study, we developed a stereo-visual hand-tracking system that enables 3D reconstruction of the trajectories of 21 anatomical landmarks of the hand during the Nine-Hole Peg Test (NHPT). Based on the reconstructed movements, kinematic parameters were computed and clinically evaluated in a cohort of 162 patients with MS by analyzing their associations with the Expanded Disability Status Scale (EDSS) and the CogEval cognitive assessment. Higher neurological disability was most strongly associated with prolonged NHPT completion time and with altered thumb movement patterns, characterized by increased movement linearity and reduced jerkiness. With respect to cognitive status, task execution speed emerged as the key contributing factor. The proposed approach is non-invasive and cost-effective, providing a detailed insight into movement quality and representing a promising tool for the objective and scalable monitoring of motor changes in MS.

1 Uvod

Nevrološke motnje, kot so multipla skleroza, Parkinsonova bolezen in cerebralna paraliza, močno vplivajo na kakovost življenja bolnikov, saj pogosto vodijo do poslabšanja motoričnih sposobnosti (Yozbatiran et al., 2006). Natančno in objektivno ocenjevanje fine motorike zgornjih okončin je ključnega pomena za spremljanje napredovanja bolezni in načrtovanje ustrezne rehabilitacije.

V klinični praksi se kot standardizirano orodje za ocenjevanje ročnih spretnosti pogosto uporablja test devetih zatičev (angl. *Nine Hole Peg Test*, NHPT) (Oxford Grice et al., 2003). Test je zaradi svoje enostavnosti in ponovljivosti široko sprejet, vendar pa je njegova glavna pomanjkljivost v tem, da pri tradicionalni izvedbi merimo zgolj skupni čas, ki ga preiskovanec potrebuje za vstavljanje in odstranjevanje vseh zatičev. Takšen pristop spregleda številne pomembne informacije o dinamiki gibanja, kot so trajektorije dlani, tresenje rok in strategije, ki jih bolnik uporabi pri izvajanju naloge.

Poleg motorične funkcije so za multiplo sklerozo značilne tudi kognitivne motnje, ki se pogosto kažejo v upočasnjem procesiranju informacij in zmanjšani pozornosti. Ker se kognitivni upad pri bolniki z MS pogosto kaže v spremenjenih vzorcih gibanja, bi lahko podrobnejša kinematična analiza NHPT služila kot boljši kazalnik bolnikovega nevrološkega stanja (Yozbatiran et al., 2006).

V dosedanjih poskusih izboljšave diagnostične vrednosti NHPT so raziskovalci preizkusili različne senzorske pristope. Digitalizirane različice NHPT z vgrajenimi senzorji so izboljšale časovno natančnost in standardizacijo testa ter pokazale visoko zanesljivost, a ostajajo omejene na izhodne meritve, kot sta trajanje testa in število napak (Mythili, 2025). Za oceno kinematičnih parametrov roke, kot sta hitrost in zveznost gibanja, so bili med izvajanjem NHPT uporabljeni tudi nosljivi sistemi na osnovi inercialnih merilnih enot (Bai et al., 2022), a ta pristop dodatno obremenjuje pacienta. Virtualni testi vstavljanja zatičev nadomestijo fizično platformo testa z vmesniki, ki zajemajo kinematične podatke, vendar na račun omejene primerljivosti z uveljavljenimi kliničnimi normami (Kanzler et al., 2020). V tem delu predlagamo video-analizo kinematike roke, ki ohranja uveljavljen potek NHPT in je povsem neinvazivna metoda.

Cilj te študije je razvoj in ovrednotenje stereo-vizualnega sistema za sledenje dlani, ki omogoča sledenje in 3D rekonstrukcijo trajektorij 21 anatomskih točk na dlani in prstih med izvajanjem NHPT. Na podlagi rekonstruiranih trajektorij je bila izvedena kvantitativna analiza gibalnih vzorcev roke z uporabo izbranih kinematičnih parametrov, kot sta skupna prepotovana pot in hitrost gibanja. Znanstveni prispevek študije je dodatno nadgrajen s klinično validacijo predlaganega pristopa na bolnikih z multiplo sklerozo (MS). Ta vključuje analizo povezav med izpeljanimi kinematičnimi meritvami ter uveljavljenimi kliničnimi ocenami motoričnih in kognitivnih funkcij, kot sta razširjena lestvica stopnje prizadetosti (angl. *Expanded Disability Status Scale*; EDSS) in kognitivna ocena CogEval, s čimer se preverjata klinična relevantnost in potencialna uporabnost metode v praksi.

2 Materiali in eksperimentalni protokol

2.1 Podatki

V študijo je bilo prvotno vključenih 186 preiskovancev. Motorična funkcija rok je bila testirana po standardiziranem protokolu NHPT, ki je obsegal dve zaporedni ponovitvi z vsako roko (dominantno in nedominantno). Pri izvedbi in vrednotenju smo sledili metodološkim usmeritvam, ki jih navajajo (Čelofiga et al., 2021) (Measurement properties of the Nine Hole Peg Test, 2026), kar zagotavlja visoko stopnjo zanesljivosti zajetih časovnih parametrov in primerljivost z uveljavljenimi kliničnimi normativi. Zaradi tehničnih omejitev smo iz nadaljnje analize izključili 19 oseb, pri katerih je delež neveljavnih 3D rekonstrukcij izvedbe testa NHPT presegal 20 %. Končni vzorec za analizo stopnje nevrološke prizadetosti je zajemal 162 udeležencev (110 žensk in 52 moških) s povprečno starostjo 47,85 let (mediana 47,0 let; razpon 23–82 let). Glede na ročnost je bilo v tej skupini 155 desničarjev in 7 levičarjev. Stopnja prizadetosti, ocenjena z lestvico EDSS, je za vključene preiskovance v povprečju znašala 1,90 (mediana 1,0; razpon 0,0–6,5), kar odraža vključenost pacientov z blago do zmerno stopnjo invalidnosti.

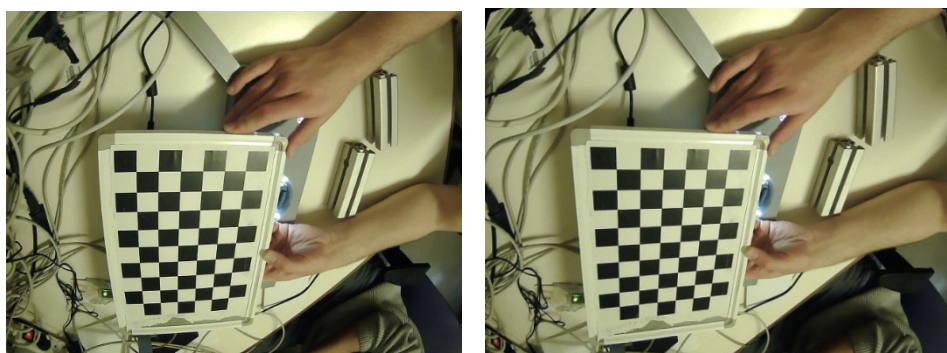
Podskupina 107 udeležencev (70 žensk in 37 moških) s povprečno starostjo 46,35 let (mediana 45,0 let; razpon 23–74 let) je bila dodatno ocenjena s kognitivnim testom CogEval. V tej skupini so bili 103 desničarji in 4 levičarji, povprečni Z-rezultat kognitivnega preizkusa pa je znašal $-0,51$ (mediana $-0,40$; razpon od $-3,36$ do $1,76$). Takšna sestava vzorca omogoča celovito analizo povezanosti med

motorično izvedbo testa NHPT, stopnjo nevrološke prizadetosti ter kognitivnim statusom preiskovancev.

2.2 Metode

2.2.1 Opis sistema NHPT in zajema videoposnetkov

Nadgrajeni sistem NHPT temelji na stereo postavitvi dveh sinhroniziranih kamer (kamere USB znamke ELP, tip USBFHD01ML28.), ki sta nameščeni nad testno ploščo na levi in desni pod kotom približno 30° glede na središče testne plošče. Video zajemata z ločljivostjo 640 x 480 pikslov in 25 slik na sekundo. Posamezno kamero smo kalibrirali z vzorcem šahovnice 9 x 6 s polji dimenzije 20×20 mm, z zajemom 15-25 slik v različnih pozah šahovnice, in določili intrinzične parametre in geometrijska popačenja leč z OpenCV (OpenCV, 2026) (slika 1).



Slika 1: Primer slike pred (levo) in po (desno) korekciji popačenja leče.

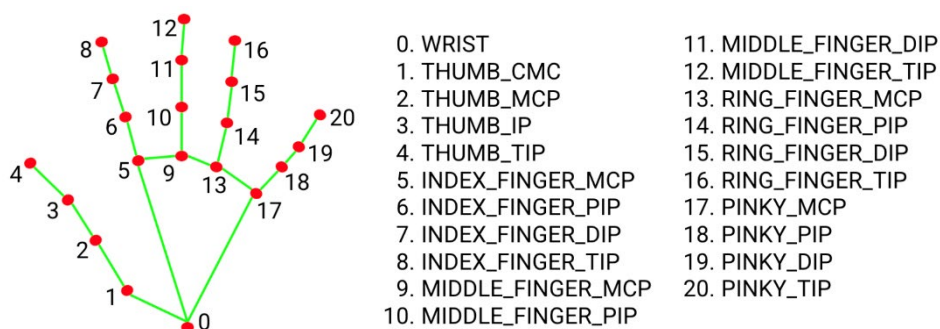
Vir: lasten.

Za kalibracijo stereo para smo na podlagi lokalizacije 18 luknjic na testni plošči in dveh dodatnih točk izven ravnine na podlagi kalibra reševali sistem enačb Longuet-Higgins. S tem smo definirali stabilen referenčni koordinatni sistem, vezan na testno platformo.

2.2.2 Analiza videoposnetkov

Kinematične parametre smo pridobili z lokalizacijo in sledenjem anatomskih točk roke, časovnim filtriranjem in 3D rekonstrukcijo. S knjižnico Python MediaPipe (google-ai-edge, 2025) smo lokalizirali 21 anatomskih točk, ki vključujejo zapestje,

členke in konice vseh petih prstov (slika 2). S triangulacijo pripadajočih parov anatomskih točk smo izvedli 3D rekonstrukcijo (`cv2.triangulatePoints` v OpenCV (OpenCV, 2026)) in tako dobili trajektorijo gibanja posamezne točke. Za glajenje trajektorij, zaradi šuma pri detekciji točk, smo uporabili filter Savitzky-Golay. Filter smo uporabili izključno na neprekinjenih veljavnih segmentih trajektorije vsakega prsta posebej, kar preprečuje popačenja na robovih manjkajočih zaznav. Glajene vrednosti smo uporabili le za izračun metrik, ki temeljijo na odvodih (hitrost, pospešek, sunkovitost, premori in stabilnost), saj so ti parametri posebej občutljivi na visokofrekvenčni šum.



Slika 2: Prikaz 21 značilnih točk dlani, kot jih zazna model MediaPipe Hands.

Vir: Hand Landmark Detection Guide (Google AI for Developers, 2026), Google AI for Developers.

Licenca: Creative Commons Attribution 4.0 (CC BY 4.0).

Za izračun prostorskih metrik in metrik »ravnosti poti« smo uporabili surove interpolirane položaje, da bi ohranili dejansko geometrijo giba. V primerih krajših intervalov brez veljavnih detekcij smo uporabili linearno interpolacijo za imputacijo, daljše prekinitev smo izločili iz nadaljnje analize.

Kinematične parametre smo strukturirali v naslednje skupine:

- **Prostorske in geometrijske značilnice:** te metrike so izračunane na surovih koordinatah in vključujejo skupno dolžino poti, volumen delovnega prostora, dolžino diagonale delovnega prostora, največji polmer dosega ter indeks planarnosti giba, ki ocenjuje, v kolikšni meri se gibanje odvija v eni ravnini. V to skupino smo dodali tudi celotno trajanje testa.

- **Ravnost in vijugavost poti:** skupina parametrov ocenjuje geometrijsko učinkovitost poti. Vključuje povprečno lokalno ravnost, deseti percentil lokalne ravnosti za zaznavo najbolj odstopajočih segmentov ter globalno vijugavost, ki je definirana kot razmerje med celotno dolžino poti in diagonalo delovnega prostora.
- **Hitrost, pospešek in sunkovitost:** dinamični parametri so izračunani na glajenih trajektorijah. Poleg povprečnih in maksimalnih vrednosti hitrosti, pospeška in sunkovitosti smo izračunali tudi koeficiente variacije, ki služijo kot indikatorji stabilnosti hitrostnega profila. Sunkovitost (angl. jerk) pri tem neposredno meri gladkost gibanja.
- **Prekinitve gibanja:** premor smo empirično definirali na ravni posameznega prsta kot segment, kjer hitrost pade pod pet odstotkov preiskovančeve maksimalne hitrosti in traja vsaj 320 milisekund. Spremljali smo število premorov, celoten čas trajanja vseh premorov, delež časa premorov glede na celotno trajanje testa ter povprečno trajanje posamezne prekinitve.
- **Stabilnost in variabilnost:** te značilnice vključujejo variabilnost hitrosti ter smerno variabilnost, ki meri disperzijo kotov vektorjev hitrosti v prostoru.
- **Koordinacija med prsti:** za oceno fine motorike smo analizirali interakcijo med palcem in kazalcem. Metrike vključujejo povprečno razdaljo med prstoma, variabilnost te razdalje in korelacijo njunih hitrosti.

Vse navedene značilnice so bile izračunane za vsako izvedbo testa posebej, kar omogoča statistično analizo povezanosti med kinematičnim profilom gibanja, stopnjo nevrološke prizadetosti in kognitivnim statusom preiskovancev.

2.2.3 Statistična analiza

Za vsakega udeleženca smo določili povprečne vrednosti kinematičnih značilnic iz vseh njegovih veljavnih testiranj. Zaradi visoke stopnje kolinearnosti med kinematičnimi parametri smo izvedli postopek krčenja nabora podatkov, pri čemer smo odstranili značilnice z močno medsebojno korelacijo, kjer je Spearmanov koeficient presegal vrednost 0,9.

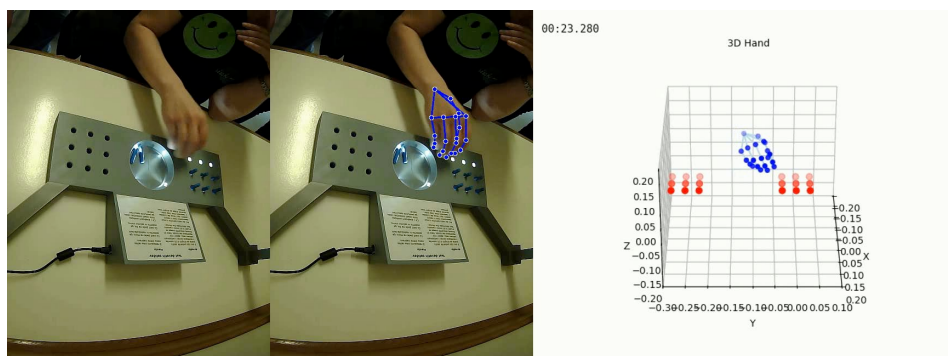
Korelacijo med izbranimi kinematičnimi značilnicami in kliničnimi ocenami smo preverjali z dvema pristopoma. Za oceno vpliva na stopnjo nevrološke prizadetosti smo uporabili ordinalno logistično regresijo, medtem ko smo kognitivni status

preverjali z linearno regresijo. V oba modela smo kot dodatni spremenljivki vključili tudi starost in spol preiskovancev. Vzporedno smo za neparametrično potrditev ugotovitev izračunali Spearmanove koeficiente korelacije med kliničnimi ocenami in kinematičnimi parametri na ravni pacienta.

3 Rezultati

3.1 Videoanaliza

Slika 3 prikazuje primer dobljene 3D rekonstrukcije. Uspešnost sledenja smo ocenili z izračunom napake reprojekcije, ki predstavlja razdaljo med dejansko zaznano točko na sliki in projekcijo rekonstruirane 3D točke nazaj na 2D ravnino. Povprečna napaka reprojekcije je znašala 7,2 mm. Za izboljšanje natančnosti smo izvedli podobnostno poravnavo (angl. similarity transformation), s katero smo 3D model optimalno skalirali, rotirali in premaknili glede na znane dimenzije testne plošče. S tem postopkom smo povprečno napako rekonstrukcije znižali na 3,1 mm.



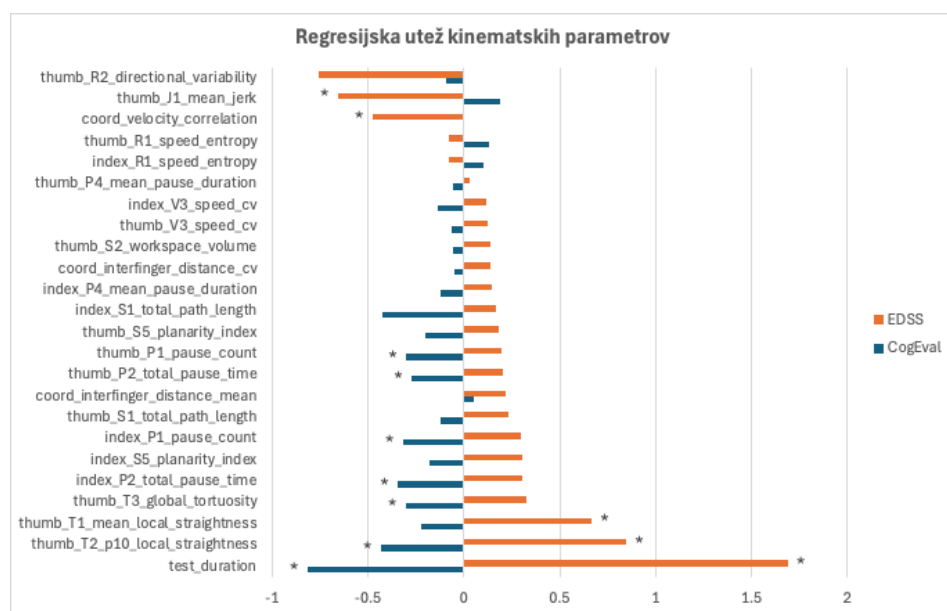
Slika 3: Izsek 3D rekonstrukcije med izvajanjem NHPT. Lokacije luknjic so označene *rdede*, anatomske točke roke pa *modro*.

Vir: lasten.

S postopkom zmanjševanja dimenzionalnosti smo začetni nabor kinematičnih parametrov omejili na 24 ključnih za nadaljnjo obdelavo.

3.2 Povezava s kliničnimi ocenami

Regresijska analiza na ravni preiskovanca, ki je vključevala prilagoditev za starost in spol, je razkrila več statistično značilnih povezav med kinematičnimi parametri in stopnjo nevrološke prizadetosti po EDSS (slika 4). Višja stopnja prizadetosti po lestvici EDSS je najmočnejše povezana s podaljšanim časom izvedbe testa. Poleg samega časa so se kot pomembni izkazali predvsem parametri gibanja palca. Ugotovili smo, da višjo stopnjo nevrološke prizadetosti spremljata visoka linearnost gibanja (*thumb_T1_mean_local_straightness*) in nižja povprečna sunkovitost gibanja palca (*thumb_J1_mean_jerk*), kar nakazuje na bolj upočasnjeno oziroma togo izvajanje gibov pri osebah z večjimi motoričnimi omejitvami.



Slika 4: Regresijske uteži kinematskih parametrov. Stolpci z oznako * so pri statistično značilnih parametrih (popravljen p-vrednost < 0.05).

Vir: lasten.

Pri kognitivnem statusu, ocenjenem s preizkusom CogEval, so statistično značilnost ohranili trajanje izvedbe testa, linearnost in vijugavost gibanja palca, pavze kazalca in palca. Rezultati potrjujejo, da so boljše kognitivne dosežke dosegali preiskovanci, ki so nalogo opravili hitreje in z neprekinjenimi direktnimi gibi. Rezultati nakazujejo, da je hitrost izvedbe v tem kontekstu primarni dejavnik uspešnosti.

Vzporedna neparametrična analiza s Spearmanovimi korelacijami (*rezultati niso prikazani*) je potrdila te povezave, ki smiselno podpirajo ugotovitve na podlagi regresijskih modelov.

4 Razprava

V tej študiji smo na podlagi videoanalize stereo para kamer opravili sledenje in 3D rekonstrukcijo anatomskih točk roke. Na kohorti 162 bolnikov MS smo pokazali, da so kinematični parametri izvedbe testa NHPT, zlasti čas izvedbe in značilnosti gibanja palca, tesno povezani s stopnjo nevrološke prizadetosti, medtem ko je kognitivni status preiskovancev predvsem odvisen od hitrosti izvedbe naloge. Ugotovitve potrjujejo, da podrobna analiza gibanja ponuja dodatne, klinično relevantne informacije o motoričnih omejitvah, ki presegajo zgolj končne časovne meritve.

Ugotovljeni gibalni vzorci, zlasti podaljšan čas izvedbe ter povečana linearnost in zmanjšana sunkovitost gibanja palca, so skladni z znanimi motnjami fine motorike pri bolnikih z multiplo sklerozo, kjer demielinizacija in motena senzomotorična integracija vodita v počasnejše, bolj nadzorovane in manj avtomatizirane gibe. Takšna strategija gibanja lahko odraža kompenzacijski mehanizem za ohranjanje natančnosti pri zmanjšani motorični zanesljivosti, hkrati pa povečuje kognitivno obremenitev med izvedbo naloge. Povezava med daljšim trajanjem testa in slabšim kognitivnim izidom dodatno podpira koncept tesne interakcije med kognicijo in motoriko pri MS, kjer upad kognitivnih funkcij omejuje učinkovitost motoričnega načrtovanja in prilagajanja, kar se izrazi v manj tekočih in bolj rigidnih gibalnih vzorcih.

V primerjavi s sorodnimi študijami, ki motorično prizadetost pri MS večinoma ocenjujejo zgolj s celotnim časom izvedbe NHPT, z nosljivimi senzorji ali novimi okolji, ta študija prinaša nov doprinos z na vizualni informaciji osnovano 3D analizo fine motorike, ki omogoča podrobno prostorsko razčlenitev gibanja prstov brez dodatnega bremena za preiskovanca. V primerjavi z nosljivimi napravami (Bai et al., 2022) je pristop neinvaziven, cenovno dostopen in ohranja protokol in s tem klinično veljavnost testa (Čelofiga, Puh, 2021), hkrati pa zagotavlja klinično relevantne informacije o kakovosti gibanja, kar predstavlja pomembno dodano vrednost za objektivno in skalabilno spremljanje motoričnih sprememb.

Zahvala

Ta študija je bila podprta s strani Javne agencije za znanstvenoraziskovalno in inovacijsko dejavnost Republike Slovenije (temeljni raziskovalni program št. P2-0232 in raziskovalni projekt št. J2-70102).

Literatura

- Assessment of upper extremity function and its relation with fatigue, cognitive function and quality of life in multiple sclerosis patients. *Journal of the Neurological Sciences*, 246(1–2), 117–122. <https://doi.org/10.1016/j.jns.2006.02.018>
- Bai, L., et al. (2022). Upper limb position tracking with a single inertial sensor using dead reckoning method with drift correction techniques. *Sensors*, 23(1), 360. <https://doi.org/10.3390/s23010360>
- Čelofiga, N., Puh, U. (2021). Measurement properties of the nine hole peg test in patients with multiple sclerosis and patients after stroke. (2026). ResearchGate. https://www.researchgate.net/publication/355958657_Merske_lastnosti_testa_devetih_zatic_ev_pri_pacientih_z_multiplo_sklerozo_in_pacientih_po_mozganski_kapi Measurement properties of the nine hole peg test in patients with multiple sclerosis and patient
- Google AI for Developers. (2026, February 1). Hand landmarks detection guide. https://ai.google.dev/edge/mediapipe/solutions/vision/hand_landmarker
- google-ai-edge. (2025, February 11). *Mediapipe (C++)*. GitHub. <https://github.com/google-ai-edge/mediapipe>
- Kanzler, C. M., Schwarz, A., Held, J. P. O., Luft, A. R., Gassert, R., & Lamberg, O. (2020). Technology-aided assessment of functionally relevant sensorimotor impairments in arm and hand of post-stroke individuals. *Journal of NeuroEngineering and Rehabilitation*, 17, 128. <https://doi.org/10.1186/s12984-020-00748-5>
- Mythili, D. (2025). Reliability and validity of the digital version of the modified standardized Nine Hole Peg Test with an autotimer. *Archives of Technical Science*, 2(33), 657–667. <https://doi.org/10.70102/afts.2025.1833.657>
- OpenCV. (2026, February 1). Open Computer Vision Library. <https://opencv.org/>
- Oxford Grice, K., Vogel, K. A., Le, V., Mitchell, A., Muniz, S., & Vollmer, M. A. (2003). Adult norms for a commercially available Nine Hole Peg Test for finger dexterity. *American Journal of Occupational Therapy*, 57(5), 570–573. <https://doi.org/10.5014/ajot.57.5.570>

PROJEKT HORIZON EUROPE ONMOVEID: TEHNIČNO OZADJE AVTOMATIZACIJE PREHODA SCHENGENSKE MEJE

KLEMEN VOVK, ROMANELA LAJČIČ, ŽAN KOGOVŠEK,
JERNEJ SABADIN, BLAŽ MEDEN, ŽIGA EMERŠIČ,
PETER PEER

Univerza v Ljubljani, Fakulteta za računalništvo in informatiko, Ljubljana, Slovenija
klemen.vovk@fri.uni-lj.si, romanela.lajcic@fri.uni-lj.si, zan.kogovsek@fri.uni-lj.si,
jernej.sabadin@fri.uni-lj.si, blaz.meden@fri.uni-lj.si, ziga.emersic@fri.uni-lj.si,
peter.peer@fri.uni-lj.si

Projekt OnMoveID si prizadeva nadomestiti stacionarne kontrole schengenskega mejnega nadzora s preverjanjem "med gibanjem" (move-through) za kopenske, morske in zračne meje. Z uporabo razširjenih denarnic za evropsko digitalno identiteto (EUDI) in standardnih pametnih telefonov želi projekt povečati pretočnost potnikov ter hkrati izboljšati varnost. Članek obravnava dve ključni komponenti 7. delovnega sklopa projekta: razvoj robustnih algoritmov za biometrično prepoznavanje in implementacijo naprednih mehanizmov za odkrivanje napadov. Te naloge naslavlja jo tehnične izzive preverjanja istovetnosti z visoko stopnjo zanesljivosti na mobilnih napravah, ki so izpostavljene tveganjem in omejitvam delovanja v nenadzorovanih okoljih.

DOI
[https://doi.org/
10.18690/um.feri.4.2026.5](https://doi.org/10.18690/um.feri.4.2026.5)

ISBN
978-961-299-111-1

Ključne besede:
OnMoveID,
schengenski mejni nadzor,
biometrično
prepoznavanje,
odkrivanje napadov,
mobilno preverjanje
istovetnosti,
preverjanje med gibanjem



Univerzitetna založba
Univerze v Mariboru

DOI
[https://doi.org/
10.18690/um.feri.4.2026.5](https://doi.org/10.18690/um.feri.4.2026.5)

ISBN
978-961-299-111-1

HORIZON PROJECT ONMOVEID: TECHNICAL BACKGROUND OF SCHENGEN BORDER CROSSING AUTOMATION

KLEMEN VOVK, ROMANELA LAJČIĆ, ŽAN KOGOVŠEK,
JERNEJ SABADIN, BLAŽ MEDEN, ŽIGA EMERŠIČ,
PETER PEER

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

klemen.vovk@fri.uni-lj.si, romanela.lajcic@fri.uni-lj.si, zan.kogovsek@fri.uni-lj.si,
jernejsabadin@fri.uni-lj.si, blaz.meden@fri.uni-lj.si, ziga.emersic@fri.uni-lj.si,
peter.peer@fri.uni-lj.si

Keywords:

OnMoveID,
schengen border control,
biometric recognition,
attack detection,
mobile identity verification,
move-through verification

The OnMoveID project aims to replace Schengen border control stationary checks with "move-through" verification for land, sea, and air borders. By using extended European Digital Identity (EUDI) wallets and standard smartphones, the project seeks to increase traveler throughput while enhancing security. This paper examines two critical components of the project's Work Package 7: the development of robust biometric recognition algorithms and the implementation of advanced attack detection mechanisms. These tasks address the technical challenges of performing high-assurance identity verification on non-trusted mobile devices in uncontrolled environments.

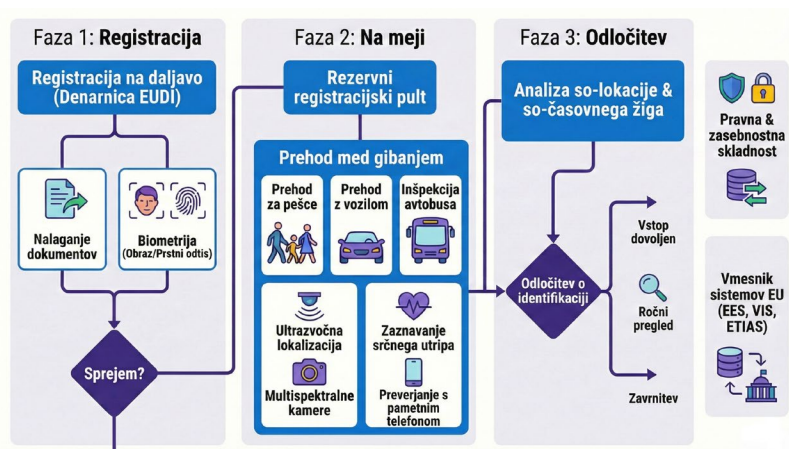


University of Maribor Press

1 Uvod

Z eksponentno rastjo števila potnikov na zunanjih mejah EU postajajo konvencionalne metode mejnega nadzora nevzdržne. Projekt OnMoveID naslavlja ta izziv s preходом na avtomatizirano preverjanje med gibanjem (glej sliko 1), ki drastično povečuje pretočnost brez kompromisov pri varnosti. V predlagani arhitekturi mobilne naprave in pametni telefoni ne delujejo kot samostojne enote za odločanje, temveč služijo primarno kot senzorski vmesniki za detekcijo obrazov in morebitni izračun osnovnih biometričnih značilk. Sama prepoznavna identitete in ujemanje pa se zaradi varnosti in računske zahtevnosti izvajata centralizirano na zavarovanih strežnikih.

Glavni inženirski in raziskovalni izziv projekta predstavlja zagotavljanje zanesljive prepoznavne osebe v izjemno raznolikih in nenadzorovanih okoljih, ki se bistveno razlikujejo od kontroliranih e-vrat (angl. e-gates). Sistem mora namreč delovati robustno ne le na letališčih, temveč tudi v kompleksnejših scenarijih na kopenskih mejah. To vključuje prepoznavo potnikov v notranjosti osebnih vozil preko zunanjih kamer ter situacije, kjer mejni policist z mobilno napravo v roki hodi skozi avtobus in sprti zajema biometrične podatke potnikov.



Slika 1: Diagram postopka prehajanja potnikov prikazuje tristopenjski sistem mejnega nadzora, ki z uporabo oddaljene registracije in biometričnih senzorjev omogoča tekoč prehod med gibanjem brez ustavljanja. Cilj je, da proces zagotavlja visoko stopnjo varnosti s preverjanjem v realnem času in neposredno povezavo z relevantnimi informacijskimi sistemi EU.

Vir: lasten.

Ta prispevek se osredotoča na specifične tehnične rešitve znotraj delovnega paketa, ki omogočajo takšno fleksibilnost. Podrobneje predstavljamo razvoj algoritmov, ki so sposobni kompenzirati spremenljive pogoje (osvetlitev, kot snemanja, premikanje policista ali vozila) ter nove pristope za zaznavo napadov, ki bi poskušali zmanipulirati zajem podatkov na terenu.

2 Razpoznavna obrazov in prstnih odtisov

Za delovanje sistema je ključen razvoj robustnega multimodalnega biometričnega cevovoda, zasnovanega za obdelavo podatkov iz heterogenih virov zajema. Čeprav sodobni biometrični sistemi dosegajo visoko zanesljivost v standardiziranih okoljih (kot so e-vrata), se pri prehodu na dinamične mejne kontrole soočamo s ključnim izzivom: variabilnostjo vhodnih podatkov, ki prihajajo iz različnih naprav.

Cilj projekta OnMoveID je tako zagotoviti zanesljivo strežniško primerjavo biometričnih značilk, ne glede na to, ali so bile te zajete s stacionarno kamero, policijsko tablico ali drugim senzorjem. Sistem pragmatično izkorišča standardne modalnosti – kamero za prepoznavo obraza in senzorje za prstne odtise – ter zagotavlja, da se ti raznoliki vnosi kljub različnim pogojem zajema uspešno preslikajo v enoten in varen proces verifikacije identitete na centralnem strežniku.

2.1 Večdelna razpoznavna obraza

Za obvladovanje variabilnosti v nenadzorovanih okoljih projekt uvaja napreden pristop, ki temelji na obdelavi video posnetkov namesto zgolj posameznih statičnih slik. Ključna prednost tega pristopa je gradnja robustnega modela identitete skozi časovno okno. Sistem iz zaporedja slik videa (ang. frames) sproti izlušči in akumulira najkvalitetnejše biometrične informacije, s čimer kompenzira trenutne motnje, kot so premikanje glave, spremenljiva razdalja ali neenakomerna osvetlitev.

Znotraj tega procesa se vsaka posamezna slika v zaporedju še dodatno analizira po segmentih (npr. oči, nos, usta), namesto da bi se obraz obravnaval le kot celota. To omogoča, da algoritem v končni model identitete selektivno vključi tiste dele obraza, ki so v določenem trenutku najbolj vidni in nezakriti. Takšna kombinacija časovne integracije in prostorske segmentacije je ključna za zanesljivost, saj empirične študije potrjujejo, da k uspešnosti razpoznavne ne prispevajo vsi deli obraza enako in je

njihova kakovost v mobilnih scenarijih spremenljiva (Lestriandoko, Veldhuis & Spreuwers, 2022)

2.2 Razpoznavanje prstnih odtisov

Čeprav je biometrija na pametnih telefonih danes vseprisotna, je njena uporaba skoraj izključno omejena na lokalno verifikacijo 1:1 (odklepanje naprave) znotraj zaprtega varnostnega ekosistema operacijskega sistema. Tu projekt naslavlja bistveno zahtevnejši inženirski izziv: omogočiti identifikacijo 1:N in preverjanje identitete v sistemih mejne kontrole, ne da bi pri tem potrebovali namensko zunanjo strojno opremo.

Glavna ovira pri tem je t.i. strojna fragmentacija. Ker razvijalci nimajo neposrednega dostopa do surovih podatkov vgrajenih kapacitivnih senzorjev (zlasti v okolju iOS), projekt uvaja rešitev, ki temelji na brezstičnem optičnem zajemu. S tem pristopom pametni telefon spremenimo v visokoločljivostni čitalec, ki ne zahteva fizičnega stika s površino.

Ključna inovacija ni le v samem zajemu, temveč v procesiranju signala. Da bi zagotovili združljivost z obstoječimi policijskimi in mejnimi bazami, ki temeljijo na kontaktnih odtisih, algoritem v realnem času ocenjuje kakovost slike po standardu NFIQ 2.0 (NIST Fingerprint Image Quality). Sistem samodejno kompenzira spremenljivo osvetlitev in perspektivna popačenja, s čimer zagotavlja skladnost s standardom ISO/IEC 29794-5. To pomeni, da lahko brezstično zajeti odtis na telefonu doseže enako raven zaupanja kot tradicionalni stacionarni čitalci, kar je predpogoj za uporabo v varnostno kritičnih procesih, kot je prestop meje. Čeprav je kakovosten zajem ključen za natančno razpoznavo, v nenadzorovanem mobilnem okolju sama natančnost ne zadostuje. Zato v naslednjem poglavju obravnavamo napade in mehanizme ugotavljanja živosti.

3 Analiza napadov in ugotavljanje živosti

Prehod na mobilno verifikacijo prinaša korenito spremembo v varnostni arhitekturi: ločitev naprave za zajem (nezaupanja vreden pametni telefon) od strežnika za odločanje. Predlog rešitve, ki ga vodi Cabinet Louis Reynaud (CLR) v sodelovanju s partnerji (vključno z UL in THALES), naslavlja tveganja te vrzeli zaupanja.

Naš pristop naslavlja kritično pomanjkljivost v obstoječih sistemih: v nenadzorovanih okoljih so prevladujoči vektorji prevare tako predstavitveni napadi (npr. tiskane fotografije, predvajanje videa, maske) kot tudi napadi z vrivanjem biometričnih podatkov, pri katerih se biometrični tok v aplikacijo vstavi programsko (npr. prek navidezne kamere, emulatorja ali namenskih orodij). Zato ugotavljanje živosti in zaznavanje vrivanja podatkov obravnavamo kot dopolnjujoča se varnostna sloja.

3.1 Zaznavanje napadov z vrivanjem podatkov

Primarni cilj tega dela je zaznavanje napadov, pri katerih napadalec uporabi navidezno kamero, emulator ali namensko programsko opremo za neposredno vstavljanje lažnih videoposnetkov v pomnilnik aplikacije, s čimer povsem obide fizično lečo kamere.

Čeprav zakonodaja (eIDAS) in okviri (PVID) že opozarjajo na to tveganje, trg ponuja malo učinkovitih rešitev. V okviru projekta razvijamo mehanizme zaznavanja, ki so skladni z novim evropskim standardom CEN/TS 18099:2024 in ciljajo na visoko raven zagotovila. Sistem analizira metapodatke senzorja in anomalije v video-toku, da zanesljivo loči med organskim posnetkom žive kamere in sintetičnim ali vrinjenim videom.

3.2 Zaznavanje predstavitvenih napadov

Za preprečevanje kraje identitete s fizičnimi pripomočki (maske, fotografije, zasloni) sistem implementira robustno ugotavljanje živosti, skladno s standardom ISO/IEC 30107-3. Namesto zastarelih aktivnih metod (zahteva po mežikanju ali nasmehu), ki jih je lažje zaobiti, se uvajajo pasivne tehnike, kot jih povzema tudi pregled sodobnih pristopov k preprečevanju predstavitvenih napadov na obrazih (Yu et al., 2023).

- **Analiza mikrogibov in teksture:** Rešitev uporablja fotopletizmografijo na daljavo (rPPG) za zaznavanje nevidnih sprememb v barvi kože, ki so posledica srčnega utripa (Liu, Lan in Yuen, 2018). Utrip je lastnost, ki je statične maske ne morejo replicirati. Hkrati modeli globokega učenja analizirajo sipanje svetlobe na površini, s čimer ločijo pravo človeško kožo od silikona, papirja ali svetlobnih pik na zaslonu (Yu et al., 2023).

- **Globina in 3D struktura:** Kjer strojna oprema to dopušča (senzorji globine ali stereo-kamere), sistem preverja 3D obliko obraza. V standardnih RGB scenarijih pa uporabljamo algoritme za strukturo iz gibanja, ki na podlagi pojava paralakse ocenijo prostorskost obraza in tako preprečijo predvajanje ploskih videoposnetkov ali fotografij (Wang et al., 2013).

3.3 Diferencialno zaznavanje napadov z morfiranjem

Napadi z morfiranjem (zlivanje dveh obrazov v eno fotografijo, kar lahko omogoča dvema osebam uporabo istega dokumenta) predstavljajo resno grožnjo mejni kontroli (Ferrara, Franco in Maltoni, 2014). Za boj proti temu je ena izmed bolj popularnih metod metoda diferencialnega zaznavanja (D-MAD): za razliko od analize ene slike, ki išče nepravilnosti na fotografiji, ta pristop izkorišča zaupanja vreden referenčni vir – sliko, varno prebrano s čipa potnega lista, in jo primerja z živo zajeto sliko (Scherhag et al., 2020). Implementacijo planiramo izvesti skladno s smernicami ISO/IEC 20059:2025 (ISO/IEC, 2025): modul primerja globoke značilke referenčne slike s čipa z živo zajeto sliko. Naši modeli bodo dodatno naučeni na naborih podatkov s transformacijami »print-scan«, kar zagotavlja odpornost tudi proti napadalcem, ki morfirano fotografijo fizično natisnejo, da bi prikrili digitalne sledi (Scherhag et al., 2020).

4 Zaključek

Razvojne aktivnosti znotraj projekta OnMoveID ne predstavljajo le programske nadgradnje, temveč temeljni inženirski odziv na omejitve trenutne mejne infrastrukture. S preходом na večdelno razpoznavo obraza in prstnih odtisov pokažemo, da je mogoče biometrično natančnost, primerljivo s stacionarnimi e-vrati, doseči tudi na komercialnih mobilnih napravah v nenadzorovanih pogojih.

Vendar pa sama biometrična natančnost brez ustrezne varnosti v odprtem ekosistemu ne zadostuje. Z implementacijo naprednih obrambnih mehanizmov, ki so skladni z najnovejšimi standardi (kot sta CEN/TS 18099 in ISO/IEC 30107-3), projekt uspešno naslavlja kritično vrzel zaupanja. Zmožnost sistema, da v realnem času loči med avtentičnim potnikom in sofisticiranim napadom s ponarejenimi podatki, je tisti ključni element, ki pametni telefon pretvori v varen potovalni žeton.

Tehnologije, predstavljene v tem članku, so pomemben korak za implementacijo vizije tekočega, brezkontaktnega prehajanja meja, ki ne sklepa kompromisov med varnostjo držav in pretočnostjo potnikov.

Viri in literatura

- Clewlow, R. R. (2016). Carsharing and sustainable travel behavior: Results from the San Francisco Bay Area. *Transport Policy*, 51, 158–164. doi:10.1016/j.tranpol.2016.01.013
- Lestriandoko, N. H., Veldhuis, R., & Spreeuwes, L. (2022). *The contribution of different face parts to deep face recognition*. *Frontiers in Computer Science*, 4, 958629. doi:10.3389/fcomp.2022.958629
- CEN. (2024). *CEN/TS 18099:2024: Biometric data injection attack detection*. European Committee for Standardization.
- ISO/IEC. (2023). ISO/IEC 30107-3:2023 Information technology — Biometric presentation attack detection — Part 3: Testing and reporting. International Organization for Standardization.
- ISO/IEC. (2025). ISO/IEC 20059:2025 Information technology — Methodologies to evaluate the resistance of biometric systems to morphing attacks. International Organization for Standardization.
- Yu, Z., Qin, Y., Li, X., Zhao, C., Lei, Z., & Zhao, G. (2023). Deep learning for face anti-spoofing: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(5), 5609–5631. <https://doi.org/10.1109/TPAMI.2022.3215850>
- Liu, S.-Q., Lan, X., & Yuen, P. C. (2018). Remote photoplethysmography correspondence feature for 3D mask face presentation attack detection. In *Proceedings of the European Conference on Computer Vision (ECCV 2018)*. https://doi.org/10.1007/978-3-030-01270-0_34
- Wang, T., Yang, J., Lei, Z., Liao, S., & Li, S. Z. (2013). Face liveness detection using 3D structure recovered from a single camera. In *Proceedings of the International Conference on Biometrics (ICB 2013)*. <https://doi.org/10.1109/ICB.2013.6612957>
- Ferrara, M., Franco, A., & Maltoni, D. (2014). The magic passport. In *Proceedings of the International Joint Conference on Biometrics (IJCB 2014)*, 1–7.
- Scherhag, U., Rathgeb, C., Merkle, J., & Busch, C. (2020). Deep face representations for differential morphing attack detection. *IEEE Transactions on Information Forensics and Security*, 15, 3625–3639. <https://doi.org/10.1109/TIFS.2020.2994750>
- ISO/IEC. (2017). ISO/IEC 29794-4:2017 Information technology — Biometric sample quality — Part 4: Finger image data. International Organization for Standardization.
- Tabassi, E., Olsen, M., Bausinger, O., Busch, C., Figlarz, A., Fiumara, G., Henniger, O., Merkle, J., Ruhland, T., Schiel, C., & Schwaiger, M. (2021). NFIQ 2: NIST Fingerprint Image Quality (NIST Interagency Report 8382). National Institute of Standards and Technology. <https://doi.org/10.6028/NIST.IR.8382>

REKONSTRUKCIJA MIŠIČNE STRUKTURE MOTORIČNIH ENOT IZ SIGNALOV HD-sEMG Z UPORABO GLOBOKIH KONVOLUCIJSKIH NEVRONSKIH MREŽ

NIKO UREMOVIĆ, NIKO LUKAČ, ALEŠ HOLOBAR

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

niko.uremovic@um.si, niko.lukac@um.si, ales.holobar@um.si

V članku predlagamo metodo za rekonstrukcijo lokacije motoričnih enot skeletnih mišic iz večkanalnih površinskih elektromiogramov (angl. High-Density Surface Electromyogram, HD-sEMG) z uporabo globokih konvolucijskih nevronske mreže. Metoda najprej uporabi dekompozicijo signala HD-sEMG v nize akcijskih potencialov motoričnih enot (angl. Motor Unit Action Potential, MUAP). Nato se uporabi globoka konvolucijska nevronska mreža, da se iz akcijskih potencialov motoričnih enot določi globina, položaj, velikost in oblika vsake motorične enote. Metoda je ovrednotena na sintetičnih podatkih mišic biceps brachii.

DOI

[https://doi.org/
10.18690/um.feri.4.2026.6](https://doi.org/10.18690/um.feri.4.2026.6)

ISBN

978-961-299-111-1

Ključne besede:

konvolucijske nevronske
mreže,
večkanalni površinski
elektromiogrami,
identifikacija motoričnih
enot,
akcijski potenciali
motoričnih enot,
dekompozicija EMG



Univerzitetna založba
Univerze v Mariboru

DOI
[https://doi.org/
10.18690/um.feri.4.2026.6](https://doi.org/10.18690/um.feri.4.2026.6)

ISBN
978-961-299-111-1

RECONSTRUCTION OF MUSCLE MOTOR UNIT STRUCTURE FROM HD-sEMG SIGNALS USING DEEP CONVOLUTIONAL NEURAL NETWORKS

NIKO UREMOVIĆ, NIKO LUKAČ, ALEŠ HOLOBAR

University of Maribor, Faculty of Electrical Engineering and Computer Science,
Maribor, Slovenia

niko.uremovic@um.si, niko.lukac@um.si, ales.holobar@um.si

Keywords:

convolutional neural
networks,
high-density surface
electromyography,
muscle motor unit
reconstruction,
motor unit action potential,
EMG decomposition

In this paper, we propose a method for identifying the anatomic properties of skeletal muscle motor units from high-density surface electromyography (HD-sEMG) signals using deep convolutional neural networks. The method first applies decomposition of the HD-sEMG signal into sequences of motor unit action potentials (MUAPs). A deep convolutional neural network is then used to estimate the depth, location, size, and shape of each motor unit from the extracted MUAPs. The method is evaluated on synthetic datasets of the biceps brachii muscle.



University of Maribor Press

1 Uvod

Razumevanje organizacije in prostorskih informacij motoričnih enot (angl. motor unit, ME) skeletnih mišic je ključno za analizo nevromišične aktivnosti, diagnostiko živčno-mišičnih obolenj ter razvoj naprednih sistemov za povratne informacije o delovanju motoričnega sistema in nevroprotektiki [1]. Motorična enota predstavlja funkcionalno enoto, sestavljeno iz motoričnega nevrona in mišičnih vlaken, ki jih inervira (Merletti in Farina, 2016). Njena prostorska porazdelitev neposredno vpliva na obliko in amplitudo izmerjenih elektromiografskih signalov (Farina idr., 2004).

Večkanalni elektromiogram (HD-sEMG) omogoča neinvazivno zajemanje prostorsko bogatih informacij o električni aktivnosti mišic (Farina idr., 2004). Sodobne metode dekompozicije, kot je večkanalna slepa ločitev virov s kompenzacijo konvolucijskih jeder (angl. Convolution Kernel Compensation, CKC), ki jo je predlagal Holobar s sodelavci (Holobar in Zazula, 2007), omogočajo zanesljivo razčlenitev signalov HD-sEMG na zaporedja akcijskih potencialov motoričnih enot (APME, angl. Motor Unit Action Potential - MUAP). Takšne metode so bistveno razširile možnosti analize nevralnega nadzora skeletnih mišic v primerjavi s klasičnimi bipolarnimi sistemi EMG (Farina idr., 2025).

Kljub napredku na področju dekompozicije ostaja odprto vprašanje identifikacije anatomske lastnosti motoričnih enot iz površinskih meritev EMG. Večina obstoječih raziskav se osredotoča na časovno dinamiko sproženja ali oceno prevodne hitrosti mišičnih vlaken (Holobar idr., 2014; Farina in Merletti, 2004; Dideriksen in Del Vecchio, 2023; Farina in Holobar, 2016), medtem ko je neposredna ocena prostorske strukture (globina, velikost, orientacija in oblika ME) bistveno manj raziskana. Prostorske metode kartiranja motoričnih enot sicer obstajajo, vendar praviloma temeljijo na heurističnih metrikah amplitude ali centroida porazdelitve signala (Cescon idr., 2007; Merletti idr., 2008; Caspo idr., 2015).

Vzporedno z razvojem HD-sEMG je globoko učenje doseglo izjemne uspehe pri modeliranju kompleksnih prostorsko-časovnih podatkov (Bao idr., 2022; Zaim idr., 2025; Deji Dere in Lee, 2023; Cote-Allard idr., 2020; Rajapriya idr., 2021). Konvolucijske nevronske mreže (CNN) so se izkazale kot učinkovite pri analizi biosignalov, vključno s signali EMG, predvsem za klasifikacijske naloge (Zhai idr., 2017; Ameri idr., 2018; Farrukh Qureshi idr., 2022). Vendar je uporaba globokih

modelov za regresijsko rekonstrukcijo fizioloških parametrov motoričnih enot iz njihovih akcijskih potencialov še vedno omejena.

V tem delu predlagamo metodo za rekonstrukcijo prostorske strukture motoričnih enot iz signalov HD-sEMG. Metoda najprej uporabi preverjeno dekompozicijsko ogrodje CKC za ekstrakcijo APME, nato pa prostorsko-časovno globoko konvolucijsko mrežo za regresijo parametrov najmanjše oklepajoče elipse, ki aproksimira presek motorične enote. Parametrizacija elipse omogoča kompaktno, interpretabilno in fiziološko smiselno predstavitev strukture ME.

Glavni izvirni prispevki članka so:

- Nov pristop za rekonstrukcijo fizične strukture motoričnih enot neposredno iz APME.
- Prostorsko-časovna konvolucijska nevronska arhitektura za regresijo parametrov MVEE.
- Sistematična evalvacija na sintetični mišici biceps brachii z nadzorovanimi fiziološkimi parametri.
- Kvantitativna primerjava z referenčnimi regresijskimi metodami.

Predlagani pristop predstavlja korak proti neinvazivni 3D rekonstrukciji mišične mikrostrukture in odpira možnosti za nadaljnje vključevanje fizikalno informiranih modelov v globoko učenje za analizo elektrofizioloških signalov.

2 Metodologija

Predlagana metoda rekonstruira strukturo mišičnih motoričnih enot z dekompozicijo signala HD-sEMG v APME, nato pa uporabi globoko konvolucijsko nevronsko mrežo (angl. convolutional neural network, CNN) za izločanje značilnic iz posameznih APME in napove globino, lokacijo, velikost in obliko vsake motorične enote.

V prvem koraku se signal HD-sEMG dekomponira na posamezne APME z uporabo večkanalne slepe ločitve izvorov (angl. Blind Source Separation, BSS) (Holobar in Farina, 2014), in sicer z uporabo kompenzacije konvolucijskih jeder (angl. Convolutional Kernel Compensation, CKC) (Holobar in Zazula, 2007). Rezultat

tega koraka je signal dimenzij ($N_{MU} \times H \times W \times T$), kjer N_{MU} označuje število razpoznanih motoričnih enot, H in W označujeta višino in širino mreže elektrod (izraženi v številu kanalov), T pa označuje število vzorcev APME.

V naslednjem koraku globoka CNN izvleče značilnice posameznih APME in rekonstruira fiziološke značilnosti motorične enote z oceno elipse, ki najbolj ustreza preseku vlaken motorične enote neposredno pod mrežo elektrod. Elipsa je bila izbrana, ker zelo verodostojno predstavlja obliko preseka večine motoričnih enot in jo je mogoče opredeliti z le nekaj parametri: središče $y_0 \in \mathbb{R}^+$, $x_0 \in \mathbb{R}$, ki ustreza globini in prečnemu odmiku središča motorične enote od središča elektrode, polosi $a \in \mathbb{R}^+$, $b \in \mathbb{R}^+$, ki opredeljujeta velikost in obliko (sploščenost) preseka motorične enote, in nazadnje kot $\theta \in [0, \pi]$, ki opredeljuje nagib preseka motorične enote.

Arhitektura globoke CNN je podrobno opredeljena v algoritmu 1 in na sliki 1 ter je povzeta v naslednjih korakih:

1. Prostorske značilnosti signala se izločijo iz vsakega okvirja z uporabo dveh 2D konvolucijskih plasti z velikostjo jedra 3×3 in plasti maksimalnega združevanja (angl. max pooling) z enako velikostjo jedra in razvlačenjem velikosti 1. Po vsaki plasti se uporabi aktivacija ReLU.
2. Izločene prostorske informacije se v vsakem časovnem koraku izravnavajo v vektor.
3. Časovne značilnosti se nato izločijo iz tako pridobljenega signala z uporabo treh 1D konvolucijskih plasti z velikostjo jedra 3. Po vsaki plasti se uporabi aktivacija ReLU.
4. Prostorsko-časovne informacije se ponovno izravnavajo v en sam vektor.
5. Trije polnopolovzani sloji se uporabijo za interpretacijo prostorsko-časovnega kodiranja v parametre elipse, ki opredeljujejo fiziološke lastnosti motorične enote vhodnega signala. Aktivacija ReLU se ponovno uporabi po vsakem sloju.

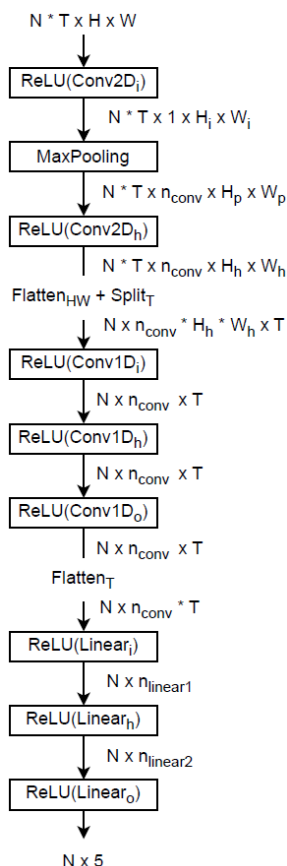
```

procedure CNN2D_INIT( $n_{feat}$ ,  $n_{conv}$ ,  $n_{linear1}$ ,  $n_{linear2}$ ,  $T$ )
   $k_{pooling} \leftarrow 3$ 
   $k_{Conv2D} \leftarrow 3$ 
   $k_{Conv1D} \leftarrow 3$ 
   $p_{pooling} \leftarrow 1$ 
   $H_i, W_i = H - k_{Conv2D} + 1, W - k_{Conv2D} + 1$ 
   $H_p, W_p = \lfloor \frac{H_i + 2 * p_{pooling} - 1}{k_{pooling}} + 1 \rfloor, \lfloor \frac{W_i + 2 * p_{pooling} - 1}{k_{pooling}} + 1 \rfloor$ 
   $H_h, W_h = H_p - k_{Conv2D} + 1, W_p - k_{Conv2D} + 1$ 
   $Conv2D_i \leftarrow Conv2D(n_{feat}, n_{conv}, (k_{Conv2D}, k_{Conv2D}))$ 
   $MaxPool2D \leftarrow MaxPool2D((k_{pooling}, k_{pooling}), p_{pooling})$ 
   $Conv2D_h \leftarrow Conv2D(n_{conv}, n_{conv}, (k_{Conv2D}, k_{Conv2D}))$ 
   $Conv1D_i \leftarrow Conv1D(H_p * W_p * n_{conv}, n_{conv}, k_{Conv1D})$ 
   $Conv1D_h \leftarrow Conv1D(n_{conv}, n_{conv}, k_{Conv1D})$ 
   $Conv1D_o \leftarrow Conv1D(n_{conv}, n_{conv}, k_{Conv1D})$ 
   $Linear_i \leftarrow Linear(n_{conv} \cdot H_h \cdot W_h \cdot T, n_{linear1})$ 
   $Linear_h \leftarrow Linear(n_{linear1}, n_{linear2})$ 
   $Linear_o \leftarrow Linear(n_{linear2}, 5)$ 

procedure FORWARD( $MUAP$ )
   $h \leftarrow ReLU(Conv2D_i(MUAP))$ 
   $h \leftarrow MaxPool2D(h)$ 
   $h \leftarrow Flatten_{HW}(ReLU(Conv2D_h(h)))$ 
   $h \leftarrow ReLU(Conv1D_i(h))$ 
   $h \leftarrow ReLU(Conv1D_h(h))$ 
   $h \leftarrow Split_{HW}(Flatten_T(ReLU(Conv1D_o(h))))$ 
   $h \leftarrow ReLU(Linear_i(h))$ 
   $h \leftarrow ReLU(Linear_h(h))$ 
   $h \leftarrow ReLU(Linear_o(h))$ 
  return  $h$ 

```

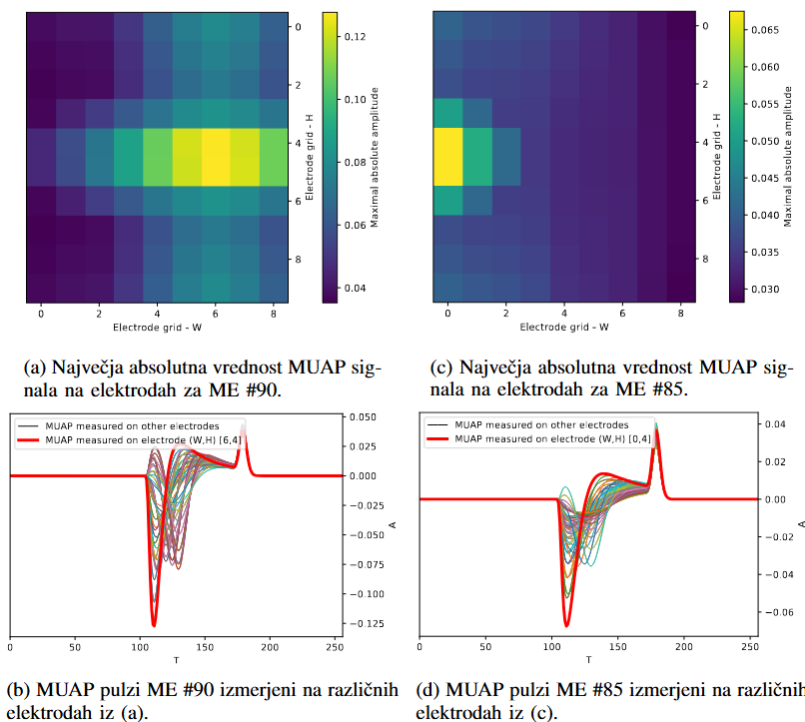
Algoritem 1: Cevovod predlagane arhitekture globoke CNN z uporabo prostorskih in časovnih konvolucijskih slojev.



Slika 1: Predlagana globoka konvolucijska nevronska arhitektura
Vir: lasten

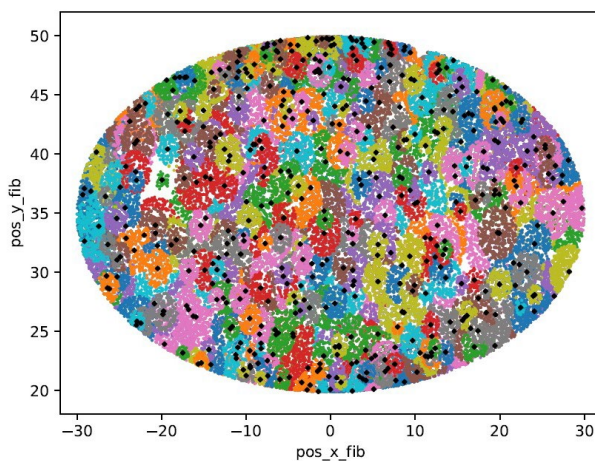
2.1 Sintetični podatki

Metoda je bila eksperimentalno ovrednotena na množici sintetičnih podatkov mišice biceps brachii. Simulacije so osnovane na modelih rekrutacije ME (Fuglevand idr., 1993) in prostorskega prevodnika (Farina in Merletti, 2001). Generiranih je bilo 10 podmnožic vzorcev različnih sintetičnih oseb, pri čemer so bile uporabljene različne fiziološke lastnosti, kot sta debelina podkožne maščobe in prevodna hitrost tkiva. Za vsako osebo so bili na mreži 10×9 elektrod z medelektrodo razdaljo 4 mm simulirani APME 500 motoričnih enot, vsak APME v dolžini 256 vzorcev. Slika 2 prikazuje podatke APME za dve večji motorični enoti.



Slika 2: Primeri APME različnih motoričnih enot.

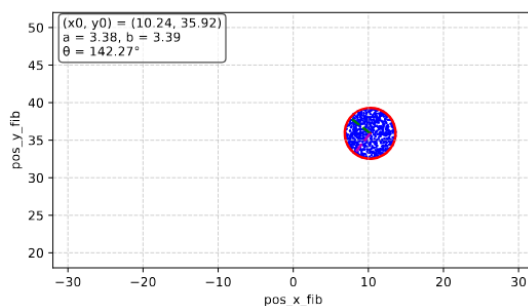
Vir: lasten



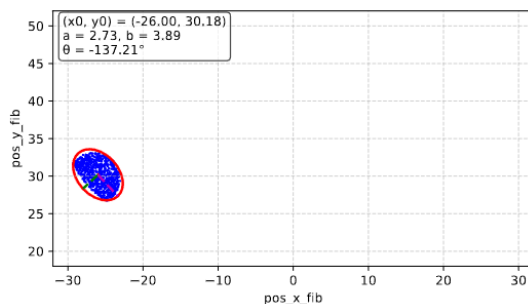
Slika 3: Struktura simulirane mišice biceps brachii, sestavljene iz 500 motoričnih enot. Vlanka motoričnih enot so predstavljena z različnimi barvami. Črni diamanti predstavljajo centre motoričnih enot.

Vir: lasten

APME vsake osebe je bil povezan s fiziološkimi značilnostmi pripadajoče motorične enote, ki so bile uporabljene v simulacijskem procesu. Pri modeliranju motoričnih enot smo se osredotočili na strukturo mišičnih vlaken. Struktura vsake motorične enote je bila opisana z 25 do 1500 vlakni. Simulirana mišična vlakna so prikazana na sliki 3 v različnih barvah, črni rombi pa označujejo središča motoričnih enot. Ker so mišična vlakna, ki tvorijo isto motorično enoto, tesno stisnjena skupaj in imajo ovalno obliko, smo se odločili, da bomo za opisovanje posameznih motoričnih enot uporabili elipse.



(a) Mišična vlakna in najmanjši oklepajoči elipsoid ene izmed večjih motoričnih enot (#90) v središču mišice.



(b) Mišična vlakna in najmanjši oklepajoči elipsoid ene izmed večjih motoričnih enot (#90) na robu mišice.

Slika 4: Primera struktur motoričnih enot, predstavljenih z mišičnimi vlakni. Izrisana je oklepajoča elipsa z najmanjšo površino (MVEE), s katero v kompaktni obliki predstavimo približno strukturo motorične enote.

Vir: lasten

Da bi natančno opredelili mišična vlakna, ki tvorijo motorično enoto, smo presek mišičnih vlaken aproksimirali tako, da smo njihova središča prilagodili v najmanjšo možno oklepajočo elipso. V splošnem se problem v poljubnem številu dimenzij

imenuje oklepajoči elipsoid z najmanjšim volumnom (angl. Minimum Volume Enclosing Ellipsoid, MVEE), za katerega približek obstaja učinkovit optimizacijski algoritem. V našem primeru gre v 2-dimenzionalnem prostoru za iskanje oklepajoče elipse z najmanjšo površino. Algoritem temelji na spustu baricentričnih koordinat, ki je opisan v (Kachiyan, 1996) in ovrednoten v (Todd in Yildirim, 2007). Slika 4 prikazuje dva primera različnih motoričnih enot, njunih mišičnih vlaken in izračunih MVEE ter njunih parametrov.

3 Rezultati

Podatkovna množica je bila razdeljena na podmnožice podatkov za učenje, validacijo in testiranje, tako da so bili izvzeti signali štirih od 10 oseb, dva za validacijo in druga dva za testiranje. Poskus je bil ponovljen v obliki 5-zložne križne validacije z različnimi subjekti v validacijskih in testnih množicah. Validacijska množica je bila uporabljena za zgodnje ustavljanje in za nalaganje uteži iz iteracije z najnižjo validacijsko izgubo po učenju. Testna množica je bila uporabljena za ovrednotenje modela po učenju. V poskusih poročamo o povprečni absolutni napaki (angl. Mean Absolute Error, MAE) napovedanih parametrov MVEE, kot tudi o povprečnem preseku nad unijo (angl. Intersection over Union, IoU) napovedanih in izračunanih MVEE, pri čemer sta obe merili povprečeni v petih iteracijah 5-zložne evalvacije.

Simuliranih 500 APME vseh 6 subjektov v naboru za učenje je bilo naključno premešano v vsaki epohi in posredovano v nevronske mreže v paketih velikosti 64. Mreža je bila učena na 3000 epohah oz. dokler se izguba na validacijski množici ni izboljšala v 500 epohah. Funkcija izgube, uporabljena v učenju, je bila srednja kvadratna napaka (angl. Mean Squared Error, MSE), uporabljena za 5 predvidenih parametrov MVEE (y_0, x_0, a, b, θ).

Rezultati analize občutljivosti so predstavljeni v tabeli 1.

Tabela 1: Primerjava uspešnosti prilagajanja parametrov MVEE pri izbiri različnih parametrizacij (števila konvolucijskih filtrov n_{conv}) globoke CNN.

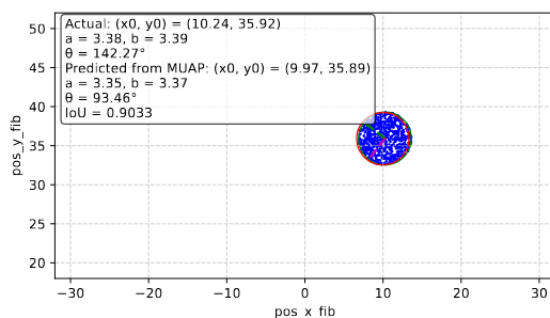
Parametrizacija	MAE (x_0, y_0, a, b, θ) [mm, mm, mm, mm, rad]	IoU
$n_{\text{conv}} = 8$	0.217, 0.227, 0.134, 0.082, 0.687	0.7970
$n_{\text{conv}} = 16$	0.201, 0.215, 0.126, 0.063 , 0.670	0.8088
$n_{\text{conv}} = 32$	0.197 , 0.206 , 0.106 , 0.066, 0.923	0.8124

Rezultati primerjave z metodami merila uspešnosti so v tabeli 2.

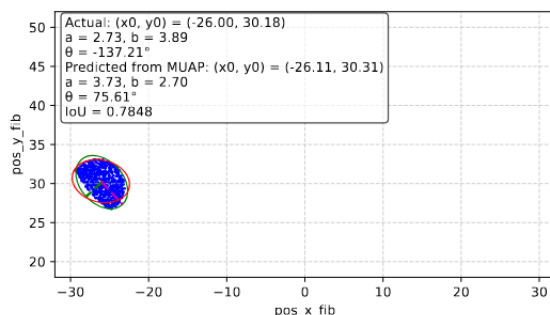
Tabela 2: Primerjava rezultatov predlagane metode s standardnimi metodami meril uspešnosti.

Metoda	MAE (x_0, y_0, a, b, θ) [mm, mm, mm, mm, rad]	IoU
RF	1.682, 1.044, 0.157, 0.097, 0.718	0.2814
MLP	2.683, 2.585, 0.305, 0.331, 0.695	0.1471
Predlagana metoda	0.197, 0.206, 0.106, 0.066, 0.923	0.8124

Na sliki 5 predstavimo napovedane MVEE za motorične enote #90 in #85 in jih primerjamo z izračunanimi MVEE iz podatkov, uporabljenih v simulaciji.



(a) Mišična vlakna motorične enote (modro), izračunan MVEE (zeleno) in napovedan MVEE (rdeče), za motorično enoto #90.



(b) Mišična vlakna motorične enote (modro), izračunan MVEE (zeleno) in napovedan MVEE (rdeče), za motorično enoto #85.

Slika 5: Primera napovedi fizične strukture motorične enote iz simuliranih APME. Narisane so izračunana oklepajoča elipsa mišičnih vlaken iz podatkov simulacije ter napovedana oz. ocenjena elipsa z uporabo predlagane metode obdelave APME.

Vir: lasten

4 Zaključek

V članku smo predstavili metodo za rekonstrukcijo prostorske strukture motoričnih enot skeletnih mišic iz signalov HD-sEMG z uporabo globokih konvolucijskih nevronske mreže. Predlagani pristop združuje preverjeno dekompozicijo signalov z metodo CKC za ekstrakcijo APME ter prostorsko-časovno CNN za regresijo parametrov najmanjše oklepajoče elipse (MVEE), ki aproksimira presek motorične enote.

Eksperimentalni rezultati na sintetičnih podatkih mišice biceps brachii kažejo, da predlagana metoda dosega bistveno boljše rezultate v primerjavi z referenčnimi regresijskimi pristopi (RF, MLP) tako v smislu povprečne absolutne napake parametrov elipse kot tudi v metriki IoU. Povprečna vrednost IoU nad 0,8 potrjuje, da je iz APME mogoče zanesljivo rekonstruirati približno geometrijsko strukturo motorične enote.

Kljub spodbudnim rezultatom predpostavljamo, da neposredna regresija parametrov elipse ni optimalna formulacija problema. Posamezni parametri (zlasti orientacija in razmerje polosi) so med seboj močno korelirani, kar lahko vodi do numerične nestabilnosti in neenakomerno porazdeljene napake v prostoru parametrov. V prihodnjem delu načrtujemo uvedbo fizikalno informiranih učnih strategij, kjer bi bila funkcija izgube neposredno definirana v geometrijskem prostoru - npr. z maksimizacijo preseka nad unijo (IoU) med napovedano in referenčno elipso oziroma z optimizacijo razdalje med implicitnimi predstavitvami elips. Takšna formulacija bi lahko izboljšala stabilnost učenja in bolje odražala dejanski cilj rekonstrukcije prostorske strukture.

Pomemben naslednji korak predstavlja validacija metode na eksperimentalno izmerjenih podatkih HD-sEMG. Prehod s sintetičnih na realne podatke bo omogočil oceno robustnosti metode ob prisotnosti šuma, neidealne geometrije elektrod ter medosebne variabilnosti.

Predlagana metoda predstavlja prvi korak proti neinvazivni rekonstrukciji mikrostrukture skeletnih mišic iz površinskih meritev EMG ter odpira nove možnosti za integracijo globokega učenja in biofizikalnega modeliranja v analizi nevro-mišičnih sistemov.

Zahvala

Avtorji se zahvaljujejo za finančno podporo Javne agencije za znanstvenoraziskovalno in inovacijsko dejavnost Republike Slovenije (št. raziskovalnega programa P2-0041).

Literatura

- Ameri, A., Akhace, M. A., Scheme, E., & Englehart, K. (2018). Real-time, simultaneous myoelectric control using a convolutional neural network. *PLOS ONE*, 13(9), e0203835.
- Bao, T., Xie, S. Q., Yang, P., Zhou, P., & Zhang, Z.-Q. (2022). Toward robust, adaptive and reliable upper-limb motion estimation using machine learning and deep learning—A survey in myoelectric control. *IEEE Journal of Biomedical and Health Informatics*, 26(8), 3822–3835.
- Cescon, C., Madeleine, P., Graven-Nielsen, T., Merletti, R., & Farina, D. (2007). Two-dimensional spatial distribution of surface mechanomyographical response to single motor unit activity. *Journal of Neuroscience Methods*, 159(1), 19–25.
- Csapo, R., Malis, V., Sinha, U., & Sinha, S. (2015). Mapping of spatial and temporal heterogeneity of plantar flexor muscle activity during isometric contraction: Correlation of velocity-encoded MRI with EMG. *Journal of Applied Physiology*, 119(5), 558–568.
- Côté-Allard, U., Campbell, E., Phinyomark, A., Laviolette, F., Gosselin, B., & Scheme, E. (2020). Interpreting deep learning features for myoelectric control: A comparison with handcrafted features. *Frontiers in Bioengineering and Biotechnology*, 8, 158.
- Dere, M. D., & Lee, B. (2023). A novel approach to surface EMG-based gesture classification using a vision transformer integrated with convolutive blind source separation. *IEEE Journal of Biomedical and Health Informatics*, 28(1), 181–192.
- Dideriksen, J. L., & Del Vecchio, A. (2023). Adaptations in motor unit properties underlying changes in recruitment, rate coding, and maximum force. *Journal of Neurophysiology*, 129(1), 235–246.
- Farina, D., & Holobar, A. (2016). Characterization of human motor units from surface EMG decomposition. *Proceedings of the IEEE*, 104(2), 353–373.
- Farina, D., & Merletti, R. (2001). A novel approach for precise simulation of the EMG signal detected by surface electrodes. *IEEE Transactions on Biomedical Engineering*, 48(6), 637–646.
- Farina, D., & Merletti, R. (2004). Methods for estimating muscle fibre conduction velocity from surface electromyographic signals. *Medical & Biological Engineering & Computing*, 42(4), 432–445.
- Farina, D., Merletti, R., & Enoka, R. M. (2004). The extraction of neural strategies from the surface EMG. *Journal of Applied Physiology*, 96(4), 1486–1495.
- Farina, D., Merletti, R., & Enoka, R. M. (2025). The extraction of neural strategies from the surface EMG: 2004–2024. *Journal of Applied Physiology*, 138(1), 121–135.
- Fuglevand, A. J., Winter, D. A., & Patla, A. E. (1993). Models of recruitment and rate coding organization in motor-unit pools. *Journal of Neurophysiology*, 70(6), 2470–2488.
- Holobar, A., & Farina, D. (2014). Blind source identification from the multichannel surface electromyogram. *Physiological Measurement*, 35(7), R143–R165.
- Holobar, A., & Zazula, D. (2007). Multichannel blind source separation using convolution kernel compensation. *IEEE Transactions on Signal Processing*, 55(9), 4487–4496.
- Holobar, A., Minetto, M. A., & Farina, D. (2014). Accurate identification of motor unit discharge patterns from high-density surface EMG and validation with a novel signal-based performance metric. *Journal of Neural Engineering*, 11(1), 016008.
- Khachiyan, L. G. (1996). Rounding of polytopes in the real number model of computation. *Mathematics of Operations Research*, 21(2), 307–320.
- Merletti, R., & Farina, D. (2016). *Surface electromyography: Physiology, engineering, and applications*. John Wiley & Sons.
- Merletti, R., Holobar, A., & Farina, D. (2008). Analysis of motor units with high-density surface electromyography. *Journal of Electromyography and Kinesiology*, 18(6), 879–890.

- Qureshi, M. F., Mushtaq, Z., Rehman, M. Z. U., & Kamavuako, E. N. (2022). Spectral image-based multiday surface electromyography classification of hand motions using CNN for human-computer interaction. *IEEE Sensors Journal*, 22(21), 20676–20683.
- Rajapriya, R., Rajeswari, K., & Thiruvengadam, S. J. (2021). Deep learning and machine learning techniques to improve hand movement classification in myoelectric control system. *Biocybernetics and Biomedical Engineering*, 41(2), 554–571.
- Todd, M. J., & Yildirim, E. A. (2007). On Khachiyan's algorithm for the computation of minimum-volume enclosing ellipsoids. *Discrete Applied Mathematics*, 155(13), 1731–1744.
- Zaim, T., Abdel-Hadi, S., Mahmoud, R., Khandakar, A., Rakhtala, S. M., & Chowdhury, M. E. H. (2025). Machine learning- and deep learning-based myoelectric control system for upper limb rehabilitation utilizing EEG and EMG signals: A systematic review. *Bioengineering*, 12(2), 144.
- Zhai, X., Jelfs, B., Chan, R. H. M., & Tin, C. (2017). Self-recalibrating surface EMG pattern recognition for neuroprosthesis control based on convolutional neural network. *Frontiers in Neuroscience*, 11, 379.

DECISION-LEVEL FUSION OF YOLOV8 AND POINTPILLARS: INITIAL FINDINGS

IVAN VRŠALOVIĆ,^{1,2} KRISTIJAN LENAC²

¹ University of Rijeka, Faculty of Informatics and Digital Technologies, Rijeka, Croatia
ivan.vrsalovic@uniri.hr, klenac@riteh.hr

² University of Rijeka, Faculty of Engineering, Rijeka, Croatia
ivan.vrsalovic@uniri.hr, klenac@riteh.hr

This paper presents a multi-modal perception system tailored for autonomous driving and safety monitoring in public parking environments, utilizing a dual-stage decision-level fusion of YOLOv8-seg and PointPillars. The architecture ensures precise occupancy monitoring and safety by integrating 2D instance segmentation with 3D LiDAR point clouds. A specialized decision fusion engine features a distance-based matching phase and a rescue phase to maintain detections during sensor occlusions. By implementing a 2.5m Euclidean threshold and a high-confidence YOLO override mechanism, the system effectively compensates for LiDAR sparsity. Experimental results on the KITTI dataset demonstrate significant reliability gains, notably increasing the F1-score for pedestrians by 10.11% and cars by 6.99%. These findings prove that the synergy of visual masks and geometric data provides a robust solution for real-time monitoring of vehicles and vulnerable road users (pedestrians and cyclists) in automated parking environments.

DOI

[https://doi.org/
10.18690/um.feri.4.2026.7](https://doi.org/10.18690/um.feri.4.2026.7)

ISBN

978-961-299-111-1

Keywords:

sensor fusion,
LiDAR,
YOLOv8-seg,
PointPillars,
KITTI dataset



Univerzitetna založba
Univerze v Mariboru

DOI
[https://doi.org/
10.18690/um.feri.4.2026.7](https://doi.org/10.18690/um.feri.4.2026.7)

ISBN
978-961-299-111-1

Ključne besede:
senzorska fuzija,
LiDAR,
YOLOv8-seg,
PointPillars,
nabor podatkov KITTI

ZDRUŽEVANJE MODELOV YOLOv8 IN POINTPILLARS NA RAVNI ODLOČANJA: ZAČETNE UGOTOVITVE

IVAN VRŠALOVIĆ,^{1,2} KRISTIJAN LENAC²

¹ Univerza v Reki, Fakulteta za informatiko in digitalno tehnologijo, Reka, Hrvaška
ivan.vrsalovic@uniri.hr

² Univerza v Reki, Tehniška fakulteta, Reka, Hrvaška
ivan.vrsalovic@uniri.hr, klenac@riteh.hr

Ta prispevek predstavlja multimodalni zaznavni sistem, prilagojen za avtonomno vožnjo in varnostni nadzor v okoljih javnih parkirišč, ki uporablja dvostopenjsko združevanje modelov YOLOv8-seg in PointPillars na ravni odločanja. Arhitektura zagotavlja natančno spremljanje zasedenosti in varnosti z integracijo 2D instančne segmentacije z oblaki 3D točk 3D LiDAR. Specializiran pogon za fuzijo odločitev (decision fusion engine) vključuje fazo umerjanja na podlagi razdalje in fazo reševanja (rescue phase) za ohranjanje zaznav med zakritji senzorjev. Z implementacijo 2,5-metrskega evklidskega praga in mehanizma za preglasitev z visoko stopnjo zaupanja YOLO (YOLO Override), sistem učinkovito kompenzira redkost podatkov LiDAR. Eksperimentalni rezultati na naboru podatkov KITTI izkazujejo znatno izboljšanje zanesljivosti, zlasti s povečanjem ocene F1 za pešce za 10,11 % in za avtomobile za 6,99 %. Te ugotovitve potrjujejo, da sinergija vizualnih mask in geometrijskih podatkov zagotavlja robustno rešitev za sprotno spremljanje vozil in ranljivih udeležencev v prometu (pešcev in kolesarjev) v okoljih avtomatiziranih parkirišč.

1 Introduction

Reliable object detection in dynamic environments remains a significant challenge for autonomous systems. While single-sensor approaches have made considerable progress, they are inherently limited by their physical properties. RGB cameras provide high-resolution semantic data but fail in precise depth estimation and low-light conditions. Conversely, LiDAR sensors provide accurate 3D spatial geometry but suffer from point cloud sparsity, particularly at long ranges or when detecting objects with low reflectivity.

This paper proposes a lightweight, yet robust multi-modal fusion framework designed to bridge these gaps at the decision level. By integrating YOLOv8-seg instance masks with PointPillars 3D proposals, the system achieves a higher level of spatial consistency than standalone models. The primary contribution is the development of a decision fusion engine, which introduces a rescue phase and a high-confidence override mechanism. These features allow the system to maintain detection integrity even when one sensor output is degraded or missing.

While this architecture is applicable to various robotics domains, it is particularly effective in autonomous driving and safety monitoring in parking environments. The system must precisely distinguish between closely parked vehicles and detect vulnerable road users (VRUs), such as pedestrians and cyclists, who are often occluded. Experimental results on the KITTI dataset show that our fusion strategy improves the **F1-score** for pedestrians by **10.11%** and for cars by **6.99%**, proving that decision-level synergy significantly enhances perception reliability.

2 Related works

The field of real-time 3D object detection has shifted significantly toward multi-modal integration to overcome the inherent limitations of single-sensor systems. A cornerstone of modern LiDAR-based detection is the PointPillars architecture (Yao et al., 2025), which revolutionized the domain by voxelizing point clouds into vertical columns, or “pillars”. This approach allows the system to utilize efficient 2D convolutional layers for feature extraction, making it highly suitable for real-time applications on edge hardware (Zhang et al., 2024). However, a primary challenge remains the “sparsity problem”; while PointPillars is highly effective for large objects

like cars, its performance often degrades when detecting humans (pedestrians) and cyclists, whose small cross-sections and lower reflectivity result in sparse LiDAR returns (Fei et al., 2020).

To mitigate these issues, recent research emphasizes late-fusion strategies that combine PointPillars with camera-based detectors such as YOLO (You Only Look Once) (Ngo et al., 2025). In a late-fusion or decision-level framework, the LiDAR and camera subsystems process data independently, with their high-level outputs merged in a final stage to ensure system redundancy (Pang et al., 2020). Recent evaluations on the KITTI dataset (Liao et al., 2022) demonstrate that integrating YOLOv8 segmentation masks provides (Malić et al., 2025) critical visual context that “boosts” the detection of human targets (Mitiu, n.d.). By applying a Probabilistic OR operation, the system can combine two relatively weak signals, such as a low-confidence LiDAR cluster and a marginal camera detection to successfully cross the detection threshold (Vora et al., 2020).

Furthermore, the development of robust override mechanisms has become central to maintaining safety in dynamic environments (Soyyigit et al., 2024). Innovations such as the Geometric Fallback mechanism allow the system to remain functional even when LiDAR data is missing or too sparse to project onto a camera mask (Ngo et al., 2025). In such cases, if the YOLO detection confidence is high, the system generates a “Pseudo-3D” bounding box based on camera priors and projection matrices. This ensures that the system maintains a high recall rate for humans and other vulnerable road users, even during sensor-specific failures or extreme sparsity (Mitiu, n.d.). Collectively, these works demonstrate that the synergy between Lidar and Camera fusion of data is essential for creating perception (Vora et al., 2020) pipelines that are both computationally efficient and resilient to environmental noise.

3 Methodology

The perception system is designed as a **late fusion** (He et al., 2023) architecture. This strategy processes data from the camera and LiDAR sensors through independent pipelines before reconciling their outputs at the decision level. This approach ensures high modularity and system reliability; if one sensor provides

sparse or degraded data, the high-level semantic or geometric information from the other sensor can maintain the detection integrity.

3.1 Spatial Synchronization and Calibration

The foundation of the fusion process is the spatial alignment between two distinct coordinate systems. Using a calibration module, the system establishes a geometric link between the 3D LiDAR space and the 2D image pixels. By applying a transformation matrix derived from calibration files, 3D points are projected onto the 2D image plane. This synchronization allows the system to accurately determine which 3D point clusters correspond to specific objects visible in the camera's field of view.

3.2 Dual-Stream Processing

The system utilizes two specialized detection streams that run in parallel:

- **The Semantic Stream:** This stream employs a YOLOv8-seg model to perform instance segmentation. It identifies object classes (such as Cars, Pedestrians, and Cyclists) and generates precise pixel-level masks. These masks are used to isolate specific LiDAR points, effectively filtering out environmental noise like the road surface or background buildings.
- **The geometric stream:** This stream processes 3D proposals generated by the PointPillars model. It provides the system with essential spatial data, including the precise 3D coordinates, object dimensions (length, width, height), and initial orientation.

3.3 The Decision Fusion Engine

The core of the methodology is the integration logic that merges the two streams. The engine calculates the **Euclidean distance** between the 3D centers of the camera-derived objects and the LiDAR proposals. If the centers are within a defined proximity threshold (2.5 meters), the detections are matched and combined.

To account for real-world sensor limitations, the engine follows two critical safety protocols:

- **The rescue phase:** In conditions where the camera may fail to see an object due to poor lighting or occlusion, the engine rescues any high-confidence LiDAR detection to ensure the object is not lost.
- **The high-confidence override:** Conversely, if the LiDAR fails to return a valid proposal due to reflective surfaces or low point density, the engine allows a high-certainty camera detection to override the result and generate a 3D box based on visual data.

3.4 Spatial Refinement and Orientation

Following the fusion of the two streams, the system performs a final geometric refinement. The point cloud associated with each fused object is processed using a clustering algorithm (DBSCAN) (Wang et al., 2019) to remove stray points that do not belong to the physical structure of the target.

To determine the final heading of the object, the system applies principal component analysis (PCA) (Shlens, 2014). By calculating the principal axes of the filtered point cluster, PCA determines the object's yaw (rotation). This mathematical refinement ensures that the final 3D bounding box is accurately aligned with the actual physical orientation of the detected vehicle or pedestrian.

4 Experimental Results and Discussion

The proposed fusion framework was evaluated on the KITTI dataset to assess its effectiveness in improving perception reliability for autonomous driving and safety monitoring in public parking environments. Performance was compared against the standalone PointPillars LiDAR baseline to quantify the benefits of the decision-level fusion strategy under realistic urban parking scenarios.

4.1 Quantitative Analysis

Evaluation focused on the three primary traffic participants: car, pedestrian, and cyclist. Precision, recall, F1-score, and mean intersection over union (mIoU) were used as the main performance indicators (Table 1). The best-performing

configuration was obtained using the probabilistic union fusion with relaxed dual-detection tolerance, which maximized gains for vulnerable road users while maintaining strong vehicle detection accuracy.

Table 1: Experiment results

<i>Method</i>	<i>Category</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>mIoU</i>
<i>PointPillars (baseline)</i>	Car	0.5953	0.7669	0.6703	0.8251
	Pedestrian	0.2981	0.6098	0.4004	0.6744
	Cyclist	0.5398	0.7376	0.6234	0.7311
<i>Proposed fusion system</i>	Car	0.7017	0.7572	0.7284	0.8255
	Pedestrian	0.4728	0.5596	0.5126	0.6765
	Cyclist	0.5983	0.5144	0.5532	0.7314

The fusion strategy significantly improved detection performance for Cars and Pedestrians. The F1-score for Cars increased from 0.6703 to 0.7284 (+0.0581), while Pedestrian detection showed the largest gain, rising from 0.4004 to 0.5126 (+0.1122). However, these improvements come with a slight trade-off between precision and recall across certain classes. For Cars, precision increased notably (0.5953 $\rightarrow$ 0.7017) while recall experienced a minor decrease (0.7669 $\rightarrow$ 0.7572), indicating a more conservative but more reliable detection behaviour. Similarly, for Pedestrians, the fusion system substantially improved precision (0.2981 $\rightarrow$ 0.4728) at the cost of a modest recall reduction (0.6098 $\rightarrow$ 0.5596), reflecting fewer false positives but slightly more missed detections.

This precision-recall trade-off suggests that the fusion framework prioritizes detection reliability and spatial consistency over exhaustive recall, which is particularly desirable in safety-critical parking scenarios where false alarms can negatively affect system trust and decision making.

4.2 Impact of Fusion Logic

The substantial improvement in pedestrian detection is primarily attributed to the rescue phase in the decision fusion engine. In parking scenarios, pedestrians frequently appear in shadows, behind vehicles, or partially occluded, which reduces camera-only confidence. However, their vertical structure remains evident in LiDAR point clouds. By preserving these LiDAR-originated detections when camera confidence drops, the system effectively reduces false negatives and improves recall for vulnerable road users.

For vehicle detection, the YOLO override mechanism played a critical role. In distant or sparsely sampled regions, LiDAR may fail to generate sufficiently dense 3D proposals. High confidence 2D masks from the camera were therefore allowed to override missing LiDAR detections, preventing loss of relevant objects and increasing overall precision. This complementary behaviour between modalities demonstrates the advantage of decision-level fusion over isolated sensor pipelines.

The cyclist category exhibited a slight decrease in F1-score compared to the LiDAR-only baseline. This is mainly due to the higher variability in cyclist geometry and motion, which occasionally leads to mismatches during the association stage. Nevertheless, the fusion system maintained comparable mIoU values, indicating that spatial localization remained consistent even when classification confidence fluctuated.

4.3 Qualitative Evaluation

Qualitative inspection of the fused outputs showed that integrating DBSCAN clustering with PCA-based orientation refinement produced more accurately aligned 3D bounding boxes while preserving detailed object contours from the camera stream. Compared to the baseline PointPillars predictions (Figure 1), which occasionally suffered from orientation drift due to sparse point clouds, the refined fusion results leveraged the precise 2D instance masks from YOLOv8 (Figure 2) to better constrain object extents and spatial placement. This complementary interaction allowed the system to maintain consistent detections even in cases of partial occlusion or low LiDAR density (Figure 3), where camera cues provided strong semantic guidance.



Figure 1: PointPillars detections

Source: Own



Figure 2: YOLOv8 detections

Source: Own



Figure 3: Fusion detections

Source: Own

As a result, the fused detections exhibited improved alignment with true object headings and boundaries, particularly for vehicles parked at non-standard angles and for pedestrians partially occluded by surrounding objects. The synergy between YOLOv8's rich visual segmentation and LiDAR-based geometric clustering proved especially beneficial in cluttered parking scenes, producing more stable and visually coherent 3D bounding boxes. Such improvements are critical for autonomous parking and valet applications, where accurate object orientation and extent directly influence path planning, obstacle avoidance, and overall maneuvering safety.

5 Conclusion and Future Work

This study presented a late-fusion framework combining YOLOv8-seg and PointPillars for 3D object detection, integrating semantic masks with LiDAR geometry to overcome the limitations of single-sensor systems. Evaluation on the KITTI Vision Benchmark Suite demonstrated notable improvements: car precision increased from 59.53% to 70.17%, while the pedestrian F1-score rose from 0.40 to

0.51, highlighting the effectiveness of the rescue phase in maintaining vulnerable road user detections. The use of DBSCAN clustering and PCA-based orientation refinement ensured 3D bounding boxes were spatially accurate and correctly aligned, enhancing reliability for autonomous parking and urban navigation.

A slight drop in recall for cyclists indicates room for improvement, but the overall increase in precision and detection stability supports practical deployment in safety-critical scenarios. Future work will focus on adaptive thresholding to better handle distant or small objects, and on refining the rescue and override mechanisms to maintain robust performance under adverse weather and low-visibility conditions. These enhancements aim to further strengthen the integration of semantic and geometric data, advancing the system toward reliable, real-time perception for autonomous driving and intelligent parking applications.

References

- Fei, J., Chen, W., Heidenreich, P., Wirges, S., & Stiller, C. (2020). SemanticVoxels: Sequential Fusion for 3D Pedestrian Detection using LiDAR Point Cloud and Semantic Segmentation. *2020 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, 185–190. <https://doi.org/10.1109/MFI49285.2020.9235240>
- He, T., Sun, P., Leng, Z., Liu, C., Angelov, D., & Tan, M. (2023). *LEF: Late-to-Early Temporal Fusion for LiDAR 3D Object Detection* (arXiv:2309.16870). arXiv. <https://doi.org/10.48550/arXiv.2309.16870>
- Liao, Y., Xie, J., & Geiger, A. (2022). *KITTI-360: A Novel Dataset and Benchmarks for Urban Scene Understanding in 2D and 3D* (arXiv:2109.13410). arXiv. <https://doi.org/10.48550/arXiv.2109.13410>
- Malić, D., Fruhwirth-Reisinger, C., Prutsch, A., Lin, W., Schuster, S., & Possegger, H. (2025). *GBlobs: Local LiDAR Geometry for Improved Sensor Placement Generalization* (arXiv:2510.18539). arXiv. <https://doi.org/10.48550/arXiv.2510.18539>
- Mitiu, M. A. (n.d.). *3D VISION OBJECT IDENTIFICATION USING YOLOv8*.
- Ngo, T. B., Ngo, L., Phi, A. V., Nguyen, T. T. H. T., Nguyen, A., Brown, J., & Perera, A. (2025). C2L3-Fusion: An Integrated 3D Object Detection Method for Autonomous Vehicles. *Sensors*, 25(9), 2688. <https://doi.org/10.3390/s25092688>
- Pang, S., Morris, D., & Radha, H. (2020). *CLOCs: Camera-LiDAR Object Candidates Fusion for 3D Object Detection* (arXiv:2009.00784). arXiv. <https://doi.org/10.48550/arXiv.2009.00784>
- Shlens, J. (2014). *A Tutorial on Principal Component Analysis* (arXiv:1404.1100). arXiv. <https://doi.org/10.48550/arXiv.1404.1100>
- Soyyigit, A., Yao, S., & Yun, H. (2024). VALO: A Versatile Anytime Framework for LiDAR-Based Object Detection Deep Neural Networks. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 43(11), 4045–4056. <https://doi.org/10.1109/TCAD.2024.3443774>
- Vora, S., Lang, A. H., Helou, B., & Beijbom, O. (2020). *PointPainting: Sequential Fusion for 3D Object Detection* (arXiv:1911.10150). arXiv. <https://doi.org/10.48550/arXiv.1911.10150>
- Wang, D., Lu, X., & Rinaldo, A. (2019). *DBSCAN: Optimal Rates For Density Based Clustering* (arXiv:1706.03113). arXiv. <https://doi.org/10.48550/arXiv.1706.03113>

- Yao, X., Liu, P., Zhou, J., Wang, Z., Fan, S., & Wang, Y. (2025). MAT-PointPillars: Enhanced PointPillars algorithm based on multi-scale attention mechanisms and transformer. *PLOS ONE*, 20(6), e0325373. <https://doi.org/10.1371/journal.pone.0325373>
- Zhang, T., Zhang, X., Liao, Z., Xia, X., & Li, Y. (2024). AS-LIO: Spatial Overlap Guided Adaptive Sliding Window LiDAR-Inertial Odometry for Aggressive FOV Variation. *2024 IEEE/RISJ International Conference on Intelligent Robots and Systems (IROS)*, 10829–10836. <https://doi.org/10.1109/IROS58592.2024.10801561>

RAZVOJ AVTONOMNEGA VPLIVNEŽA Z UPORABO ORODIJ UMETNE INTELIGENCE

JON ROJNIK GORŠEK, BORUT BATAGELJ

Univerza v Ljubljani, Fakulteta za računalništvo in informatiko, Ljubljana, Slovenija
jon.r.gorsek@gmail.com, borut.batagelj@fri.uni-lj.si

Vplivneži imajo pomembno vlogo pri oblikovanju javnega mnenja ter pri komuniciranju vsebin na družbenih omrežjih, kar odpira nove priložnosti za uporabo umetne inteligence pri avtomatiziranem ustvarjanju digitalnih identitet. V prispevku je predstavljen razvoj avtonomnega virtualnega vplivneža, namenjenega promociji Fakultete za računalništvo in informatiko, Univerze v Ljubljani (FRI UL), ki temelji na uporabi sodobnih modelov za generiranje besedilnih in slikovnih vsebin. Sistem samodejno zbira aktualne informacije iz izbranih spletnih virov ter z orodji umetne inteligence, oblikuje objave v slogu profesorja, ki jih nato avtomatizirano distribuira na družbena omrežja. Opisan je celoten proces zasnove, implementacije in testiranja delovanja avtonomnega agenta. Poseben poudarek je namenjen tehnološkim izzivom, priložnostim ter etičnim vprašanjem, ki spremljajo uporabo umetne inteligence pri ustvarjanju in upravljanju digitalnih osebnosti.

DOI

[https://doi.org/
10.18690/um.feri.4.2026.8](https://doi.org/10.18690/um.feri.4.2026.8)

ISBN

978-961-299-111-1

Ključne besede:

umetna inteligenca,
UI-vplivnež,
UI-agent,
družbena omrežja,
avtomatizacija

Prispevek temelji na:

Rojnik Goršek, J. (2025).

*Razvoj avtonomnega
inflencerja z uporabo orodij
umetne inteligence* diplomsko
delo, Univerza v Ljubljani,
Fakulteta za računalništvo
in informatiko, Ljubljana.



Univerzitetna založba
Univerze v Mariboru

DOI
[https://doi.org/
10.18690/um.feri.4.2026.8](https://doi.org/10.18690/um.feri.4.2026.8)

ISBN
978-961-299-111-1

Keywords:

artificial intelligence,
AI influencer,
AI agent,
social media,
automation

The paper is based on:

Rojnik Goršek, J. (2025).
*Development of an Autonomous
Influencer Using Artificial
Intelligence Tools*: bachelor's
thesis, University of
Ljubljana, Faculty of
Computer and Information
Science, Ljubljana.

DEVELOPMENT OF AN AUTONOMOUS INFLUENCER USING ARTIFICIAL INTELLIGENCE TOOLS

JON ROJNIK GORŠEK, BORUT BATAGELJ

University of Ljubljana, Faculty of Computer and Information Science,
Ljubljana, Slovenia
jon.r.gorsek@gmail.com, borut.batagelj@fri.uni-lj.si

Influencers play an important role in shaping public opinion and communicating content on social media, which opens new opportunities for the use of artificial intelligence in automating the creation of digital identities. This paper presents the development of an autonomous virtual influencer designed to promote the Faculty of Computer and Information Science at the University of Ljubljana (FRI UL), based on the use of modern AI models for generating textual and visual content. The system automatically gathers up-to-date information from selected online sources and, using AI tools, generates professor-style posts that are then automatically distributed across social media platforms. The complete process of designing, implementing and testing the autonomous agent is described. Special emphasis is placed on the technological challenges, opportunities and ethical considerations that accompany the use of artificial intelligence in creating and managing digital personas.



University of Maribor Press

1 Uvod

V tem prispevku predstavljamo razvoj in implementacijo prototipa avtonomnega virtualnega vplivneža, ki smo ga razvili za promocijo Fakultete za računalništvo in informatiko Univerze v Ljubljani (FRI UL). Namen projekta je bil raziskati, v kolikšni meri lahko umetna inteligenca samostojno ustvarja, oblikuje in objavlja vsebine na družbenih omrežjih, brez neposrednega človeškega posredovanja. Z uporabo naprednih orodij umetne inteligence (UI) in avtomatizacije smo oblikovali avtonomnega UI agenta, ki vsakodnevno poišče nove vsebine, pripravi besedilni opis, ustvari vizualni element, odgovarja na komentarje ter samodejno objavlja rezultate.

1.1 Koncept avtonomnega vplivneža

Virtualni vplivneži so digitalne osebnosti, ki nastopajo na družbenih omrežjih podobno kot človeški vplivneži, vendar so v celoti ustvarjeni s pomočjo umetne inteligence in naprednih grafičnih orodij. So lahko fotorealistični ali stilizirani in pogosto uporabljajo kompleksne modele za generiranje slik ter naravnega jezika. Zaradi te zasnove delujejo brez časovnih, prostorskih ali fizičnih omejitev, kar blagovnim znamkam in institucijam omogoča neprekinjeno komunikacijo ter popoln nadzor nad identiteto in sporočili vplivneža.

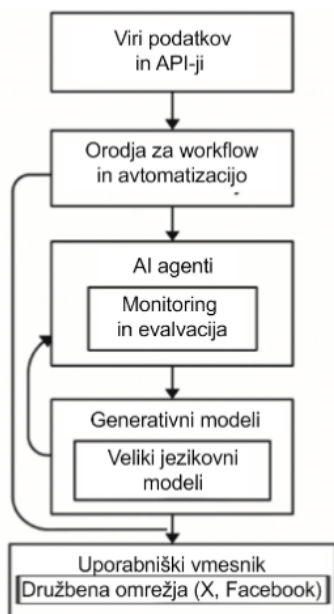
Virtualni vplivneži pomenijo tudi priložnost za eksperimentiranje z identitetami in komunikacijskimi slogi, ki bi bili v resničnem svetu omejeni. Kljub temu se s pojavom takšnih likov porajajo pomembna vprašanja, povezana z etiko, transparentnostjo in vplivom na uporabnike. Raziskave poudarjajo, da je jasna označitev virtualnih vplivnežev kot umetnih oseb ključna za ohranjanje zaupanja javnosti (Jin SV, 2019). Brez transparentnosti lahko sledilci občutijo zavajanje, kar zmanjšuje verodostojnost komunikacije. Poleg tega lahko virtualni vplivneži utrjujejo družbene stereotipe in predsodke, na kar opozarjajo strokovnjaki (Glover, 2025). Na primer, opazili so, da so znani UI vplivneži pogosto predstavljeni kot mlade, vitke ženske, kar lahko krepi idealizirane podobe in predsodke v družbi. Zato razvoj takšnih sistemov zahteva previdnost pri načrtovanju podobe vplivneža in skrbno obravnavo etičnih vidikov.

2 Arhitektura sistema

Za razvoj avtonomnega virtualnega vplivneža je bilo ključno oblikovati modularen sistem, ki omogoča zanesljivo delovanje, razširljivost in nadzor nad celotnim procesom ustvarjanja ter objavljanja vsebin. Arhitektura temelji na medsebojno povezanih komponentah, pri čemer vsak modul opravlja specifično funkcijo - od zbiranja informacij do končne distribucije objav na družbenih omrežjih. Takšna zasnova omogoča dinamično prilagajanje vsebin aktualnim dogodkom, hkrati pa zagotavlja sledljivost in avtomatizacijo, ki sta nujni za neprekinjeno digitalno prisotnost. Ključni moduli sistema so:

- Zajem vsebine: Modul samodejno preverja vire informacij - spletne strani fakultete, repozitorije diplomskih nalog (COBISS) in druge akademske vire. Če najde novo novico ali delo, ga prenese v bazo.
- Obdelava podatkov: Pridobljene vsebine se analizirajo in filtrirajo glede na datum ter unikatnost. Uporabljamo vektorske predstavitve besedila za primerjavo z že objavljenimi vsebinami, s čimer preprečimo ponavljanje.
- Generiranje objav: Glavni del sistema temelji na velikem jezikovnem modelu GPT-4, ki ustvari kratko in privlačno objavo v slovenščini. Dodamo navodila za ton komunikacije, dolžino objave in vključitev ključnikov #fri ter enega po izbiri.
- Generiranje slik: Če je potrebno, sistem aktivira model DALL-E 3, ki ustvari ilustracijo na podlagi besedilnega opisa. Tako zagotovimo, da ima vsaka objava svoj vizualni element, neodvisen od pravic uporabe.
- Objava na družbenih omrežjih: Po pripravi vsebine sistem samodejno objavi prispevek na platformah Facebook, X (Twitter), LinkedIn in Threads prek programskih vmesnikov (API).
- Avtomatsko komentiranje: Po objavi lahko sistem samodejno generira kratke, vsebinsko ustrezne komentarje kot odziv na interakcije uporabnikov.

Vsi ti moduli hranijo podatke, slike in metapodatke v centralizirani podatkovni zbirki Supabase, kar zagotavlja preglednost in omogoča naknadno analizo delovanja sistema.



Slika 3: Povezava med ključnimi orodji in ogrodji

Vir: lasten

3 Način delovanja

Dnevno sistem samodejno zažene proces ustvarjanja objav. Najprej preveri aktualne novice. Če jih najde, izbere eno in jo uporabi kot osnovo. Če novic ni, preveri nove diplomske naloge ali znanstvene članke. V primeru popolnega pomanjkanja vsebin uporabi zbirko vnaprej pripravljenih tem, ki jih imenujemo »fallback teme«. Na ta način sistem zagotavlja redno aktivnost in dosledno objavljanje. Po izbiri vsebine veliki jezikovni model (GPT-4) generira besedilo. Na primer, če gre za novico o dogodku, model pripravi vabilo ali povzetek v dveh stavkih. V primeru raziskovalnega članka pripravi kratko razlago v poljudnem jeziku. Besedilo je zasnovano tako, da vzbuja zanimanje, vključuje povezavo na vir ter ključnike.

Za vizualni element agent vpraša jezikovni model, ali bi bila slika dodatek. Če model to odobri, predlaga motiv na podlagi teme objave (npr. računalnik v laboratoriju, skupina študentov v diskusiji). Nato aktiviramo generativni model (DALLE-3), ki ustvari ilustracijo po predlogu. Tako vsaka objava dobi še ustrezno sliko, ki dopolnjuje sporočilo besedila (slika 2).

Nazadnje sistem prek integriranih vmesnikov API objavi prispevek na izbranih družbenih omrežjih. Po objavi zapis v bazi podatkov beleži rezultate (čas objave, platformo, ID objave). V določenih primerih sistem bere komentarje in jezikovni model uporabi za oblikovanje hitrega odgovora, s čimer omogočimo omejeno interakcijo s sledilci.



Slika 4: Prikaz avtomatsko ustvarjene objave s sliko, besedilom in komentarjem
Vir: lasten

4 Rezultati delovanja

Projektni sistem smo preizkušali več tednov. Ugotovili smo, da deluje stabilno in skladno z načrtom: vsak dan je bila objavljena nova vsebina, profili pa so ohranili stalno aktivnost brez potrebe po ročnem vnašanju objav. Tudi v dneh brez novih novic ali diplomskih nalog so »fallback objave« preprečevale zatišje, s čimer je profil ostal aktiven in zanimiv za sledilce. S tem smo dosegli konsistentno prisotnost na družbenih omrežjih, kar lahko pozitivno vpliva na vidnost objav (aktivni profili običajno dosežejo več sledilcev). Analiza generiranih vsebin je pokazala, da so besedila večinoma primerna, kratka in jedrnata. Jezikovni model GPT-4 se je izkazal za zelo zmogljivega; njegove prirejene objave so slovnično pravilne ter slogovno enotne, kar ustreza pričakovanim komunikacijskim standardom fakultete. Ton objav je prijazen in strokoven, kot če bi jih pisal predstavnik univerze. Slike, ustvarjene z DALL-E 3, se sicer ne uvrščajo med najkompleksnejše generativne umetnine, a učinkovito dopolnjujejo besedilo objav. Vsaka slika pritegne pozornost in hkrati zagotavlja, da so vizualni elementi avtorsko varni. S celovito zasnovo sistema smo v bistvu dobili svojega ustvarjalca UI vsebin, ki za univerzo krepi spletno prisotnost. Virtualni vplivnež deluje podobno kot pravi predstavnik znamke: dosledno objavlja relevantne informacije (dogodke, dosežke študentov, izobraževalne zanimivosti) in s tem ohranja pozornost javnosti.

Opazili smo, da je robot večino časa samostojno reševal manjše težave. Na primer, ko je GPT-4 v besedilih večkrat ponavljal monotono frazo »preberite več na povezavi«, smo navodila prilagodili, da se povabilo h kliku pojavi na drugačen način. Podobno smo zagotovili, da ključnik #fri vedno ostane na koncu objave. Zaradi teh prilagoditev so končne objave še kakovostnejše. Kljub temu smo zaznali omejitve, in sicer da model pogosto generira zelo splošne slike. V večini primerov so bile ilustracije zadostne, a občasno se je pojavila kakšna abstraktna podoba, ki ni najboljše sovpadala z besedilom. To je inherentna omejitev generativnih modelov - nimamo popolnega nadzora nad končnim izhodom. Ena izmed prihodnjih možnosti izboljšav je vključitev še naprednejših modelov (na primer GPT-5 za besedila ali GPT-Image-1 za slike), ki obljublajo večjo ustvarjalnost in skladnost med besedilom in vizualizacijo.

Med izvajanjem sistema nismo beležili večjih neposrednih odzivov uporabnikov, saj so bile vse objave na začetku objavljene prek novoustanovljenih (testnih) računov. Torej nismo pridobili značilnih všečkov, delitev ali komentarjev, ki bi omogočali

merjenje dejanske angažiranosti občinstva. Ugotavljamo, da je to omejitev testa - sistem sicer tehnično deluje in generira vsebine, vendar za resnični vpliv na družbenih omrežjih potrebuje aktivno bazo sledilcev. V praksi bi bilo potrebno ustvariti zanimanje za profil (na primer s promocijo na drugih kanalih), da bi objave dosegale širše občinstvo. Šele s primerno publiko bi lahko natančneje ocenili dosege in odziv ter na podlagi teh podatkov dodatno prilagajali slog objav. V prihodnosti načrtujemo implementacijo avtomatske analize angažiranosti (všečki, delitve, komentarji) in prilagajanje komunikacijskega tona glede na odziv občinstva.

5 Etika in prihodnji razvoj

Delovanje sistema v vlogi virtualnega vplivneža odpira več etičnih vprašanj, predvsem glede transparentnosti. Pomembno je, da je jasno označeno, da gre za digitalno ustvarjeno osebnost, saj lahko pomanjkanje takšne informacije zmanjša zaupanje sledilcev in povzroči dvome o pristnosti objav. Uporabniki morajo razumeti, da z likom komunicirajo posredno prek umetne inteligence, kar je ključno za pošteno predstavitev informacij. Raziskava opozarja, da lahko povsem umetna podoba vplivneža povzroči občutek neavtentičnosti (Jin, 2025). Sledilci se lahko počutijo prevarane, ko ugotovijo, da za profilom ni človeškega avtorja, kar posebej vpliva na mlajše občinstvo. Zato je pri zasnovi pomembno ohraniti primeren ton komunikacije in uravnoteženo »človeškost«, da se izognemo občutku hladne umetnosti. Pri oblikovanju takšnih sistemov je treba upoštevati tudi vprašanje stereotipov. Virtualni vplivneži lahko nehote reproducirajo idealizirane podobe ali pristranskosti, kar vpliva na dožemanje družbenih vlog. Zato je smiselno uporabljati nevtralne in raznolike vizualne elemente ter sproti preverjati morebitne pristranskosti modelov.

Kljub izzivom pa ostaja prihodnost virtualnih vplivnežev obetavna. Ob odgovorni uporabi lahko predstavljajo učinkovit način za ohranjanje stalne prisotnosti na družbenih omrežjih in dopolnitev človeške kreativnosti. Če so zasnovani pregledno in premišljeno, lahko pomembno prispevajo k sodobni digitalni komunikaciji.

5.1 Sklep

Z razvitim sistemom smo demonstrirali, da je mogoče ustvariti popolnoma avtonomnega virtualnega vplivneža, ki samostojno išče informacije, generira besedila in slike ter jih deli na družbenih omrežjih. Tako smo potrdili praktično uporabnost

sodobnih orodij umetne inteligence pri digitalnem komuniciranju in promociji akademskih institucij. Avtonomni vplivnež predstavlja presečišče več področij - od računalništva do družbenih ved, kar odpira številna vprašanja za prihodnje raziskave. V nadaljevanju bomo raziskali možnosti za integracijo bolj »personalizirane« interakcije z občinstvom (npr. prek video-vsebin) ter razvili metode ocenjevanja družbenega vpliva takega sistema. Na podlagi razvoja umetne inteligence in sodobnih digitalnih medijev lahko pričakujemo, da bodo virtualni vplivneži kmalu postali sestavni del komunikacije, če bomo pri tem dosledno upoštevali tehnična in etična načela.

Viri in literatura

- Jin SV, Muqaddam A, Ryu E (2019), Instafamous and social media influencer marketing. *Marketing Intelligence & Planning*, Vol. 37, No. 5, pp. 567-579, doi: 10.1108/MIP-09-2018-0375
- Glover, E. (2025). AI Influencers, Explained. *Built In*, Dostopno na: <https://builtin.com/articles/ai-influencer> (datum dostopa: 05.05.2025).
- Jin, S.V., Viswanathan, V. Threatened and empty selves following AI-based virtual influencers: comparison between followers and non-followers of virtual influencers in AI-driven digital marketing. *AI & Soc* 40, 117–131 (2025). doi: 10.1007/s00146-023-01832-9

PERSONALIZIRAN PRIKAZ DIGITALNIH SLIK ZA RAZLIČNE VRSTE BARVNE SLEPOTE

ŽIGA VINČEC, BOŽIDAR POTOČNIK, MARTIN ŠAVC

Univerza v Mariboru, Fakulteta za elektrotehniko, računalništvo in informatiko,
Maribor, Slovenija
ziga.vincec@student.um.si, bozidar.potocnik@um.si, martin.savc@um.si

Barvna slepota pomeni nezmožnost razlikovanja določenih barv ali odtenkov, ki jih ljudje z normalnim barvnim vidom zlahka prepoznajo. V članku je predstavljena programska rešitev, ki barvno slepim omogoča personalizirano prilagoditev digitalnih slik in tako olajša njihove vsakodnevne izzive. Namen rešitve je uporabniku omogočiti razumevanje vizualnega konteksta ob čim manjšem posegu v izvirno sliko. Rešitev vsebuje hiter test, s katerim določimo, katere barve so za uporabnika konfuzne. Prebarvanje je vsebinsko odvisno. Iz slike izluščimo reprezentativne barvne centre in na podlagi testa barvne slepote prepoznamo barvne pare, ki jih uporabnik težko razlikuje. Za konfuzne barvne centre z optimizacijskim postopkom poiščemo nadomestne barve, ki najmanj spremenijo videz slike, tako da se ti segmenti prilagodijo in postanejo uporabniku jasno razločljivi. Dobljeni rezultati so obetajoči.

DOI

[https://doi.org/
10.18690/um.feri.4.2026.9](https://doi.org/10.18690/um.feri.4.2026.9)

ISBN

978-961-299-111-1

Ključne besede:

barvna slepota,
daltonizacija,
personalizacija,
prebarvanje digitalnih slik,
programska oprema

Prispevek temelji na:

Vinčec, Ž. (2025).
*Personaliziran prikaz
digitalnih slik za različne vrste
barvne slepote z metodo
daltonizacije: 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.4.2026.9](https://doi.org/10.18690/um.feri.4.2026.9)

ISBN
978-961-299-111-1

Keywords:

colour vision deficiency,
daltonization method,
personalisation,
digital image recolouring,
software

The paper is based on:
Vinčec, Ž. (2025).

*Personaliziran prikaz digitalnih
slik za različne vrste barne
slepote z metodo daltonizacije;*
*master's thesis, University of
Maribor, Faculty of
Electrical Engineering and
Computer Science, Maribor.*

PERSONALIZED DIGITAL IMAGE DISPLAY FOR VARIOUS TYPES OF COLOUR VISION DEFICIENCIES

ŽIGA VINČEC, BOŽIDAR POTOČNIK, MARTIN ŠAVC

University of Maribor, Faculty of electrical engineering and computer science ,
Maribor, Slovenia
ziga.vincec@student.um.si, bozidar.potocnik@um.si, martin.savc@um.si

Colour vision deficiency is the inability to distinguish certain colours or shades that people with normal colour vision can easily recognize. This article presents a software solution that allows individuals with colour vision deficiency to personalize digital images, thereby facilitating their daily challenges. The purpose of the solution is to enable users to understand the visual context with minimal interference to the original image. The solution includes a quick test to determine which colours are confusing to the user. The recolouring is content-dependent. We extract representative colour centres from the image and, based on the colour blindness test, identify colour pairs that are difficult for the user to distinguish. For confusing colour centres, we use an optimization process to find replacement colours that change the appearance of the image as little as possible and at the same time adjust the confusing segments so that they become clearly distinguishable to the user. The results obtained are promising.



University of Maribor Press

1 Uvod

Daltonizem oziroma barvna slepota je dedna in neozdravljiva motnja zaznavanja barv, ki se med posamezniki močno razlikuje. Lahko se kaže kot blaga težava pri razlikovanju nekaterih barvnih odtenkov ali kot popolna barvna slepota, pri kateri oseba vidi le sivinske odtenke (Fairchild, 2005). Ker barve v digitalnih vsebinah pogosto nosijo ključne informacije, lahko neprilagojene barvne sheme barvno slepim otežijo ali onemogočijo razumevanje podatkov. To je posebej izrazito pri zemljevidih, grafih in signalizaciji, kjer napačna izbira barv povzroča zmedo ali celo napačno interpretacijo – še posebej pri rdeče-zelenih kombinacijah, ki so za najpogostejšo obliko barvne slepote najbolj problematične. Večina barvno slepih ne pozna natančne vrste svoje motnje, zato morajo rešitve za prilagajanje digitalnih vsebin vključevati test, ki to zanesljivo določi. V članku predstavljamo pristop za personalizirano prilagajanje digitalnih slik, ki temelji na izokromatskih tablicah. Z njimi določimo osem diskriminatornih limit, kjer barve preidejo iz neločljivih v ločljive. Iz teh limit nato oblikujemo personaliziran globalni elipsoid, s katerim opredelimo barve, ki so za uporabnika konfuzne – ne glede na vrsto barvne slepote, zaslon ali vpliv okolja.

Postopek prebarvanja slike poteka v štirih korakih. Najprej z optimizirano različico algoritma Mean Shift določimo reprezentativni nabor barv, ki najbolje opiše barvno paleto slike. Nato s pomočjo personaliziranega elipsoida, izpeljanega iz rezultatov testa, prepoznamo konfuzne barvne centre. Med njimi izberemo tiste, katerih sprememba najmanj vpliva na videz slike, in jim z optimizacijskim postopkom poiščemo nadomestne barve, ki minimalno odstopajo od izvirnih. Tako omogočimo, da barvno slepi barve jasno razlikujejo, hkrati pa ohranimo prepoznavnost slike. V zadnjem koraku vsakemu pikslu nove slike določimo ustrezno barvo: piksli nekonfuznih centrov ohranijo izvirne barve, medtem ko piksle optimiziranih centrov obarvamo glede na njihovo razdaljo od centra. S tem se izognemo nenadnim barvnim prehodom in ohranimo naraven videz slike.

2 Medicinsko ozadje

Povprečen človek zaznava svetlobo med 400 in 700 nm, kar omogoča prepoznavanje tisočih barv. Barvni vid temelji na treh vrstah čepnic – S (angl. short), M (angl. medium) in L (angl. long) – ki so občutljive na različna območja valovnih

dolžin. Njihovi signali se najprej združijo v splošni signal svetlosti, nato pa primerjajo, kar ustvari dva nasprotna barvna kanala: rdeče-zelenega (primerjava čepnic L in M) ter rumeno-modrega (primerjava kombinacije M+L s signalom čepnic S) (Nathans, 1999). Barvni vid ni enak pri vseh ljudeh, saj so pogoste različne oblike barvne slepote (glej tabelo 1). Pri najhujši, monokromatiji paličic, so vse tri vrste čepnic močno okvarjene ali nefunkcionalne, zato oseba vidi le sivinske odtenke. Podobno je pri monokromatiji čepnic, kjer deluje le ena vrsta čepnic, kar pomeni zaznavo le ene barvne osi. Obe obliki sta izjemno redki, zato je njihova pogostost težko natančno določljiva. Dikromatija je blažja oblika barvne slepote, pri kateri zaradi odsotnosti pigmenta v eni vrsti čepnic barvni vid deluje, a je bistveno okrnjen. Poznamo tri vrste: tritanopijo (nefunkcionalne čepnice S), pri kateri so težavni modro-rumeni odtenki, ter deuteranopijo in protanopijo (okvarjene čepnice M oziroma L), kjer odpove rdeče-zeleni barvni kanal (Simunovic, 2009). Deuteranopija in protanopija sta med najpogostejšimi oblikami barvne slepote in prizadeneta nekaj več kot 2 % ljudi.

Anomalna trikromatija je najpogostejša oblika barvne slepote, saj jo ima okoli tri četrtine barvno slepih. Vse tri vrste čepnic delujejo, vendar je ena spektralno premaknjena, zato je razlikovanje barv oslABLJENO — od skoraj normalnega vida do stanj, ki se približajo dikromatiji (Hunt, 2011). Poznamo tri oblike: tritanomalijo (oslABLJENE čepnice S), protanomalijo (premik čepnic L proti nižjim valovnim dolžinam) in deuteranomalijo (premik čepnic M proti višjim). Večje kot je prekrivanje krivulj, več barvnih odtenkov postane konfuznih.

Tabela 1: Statistika pojavitve vrst barvne slepote (vir: (Hunt, 2011))

Tip	Pogostost pojavitve (%)	
	Moški	Ženske
Deuteranomalija	4,9	0,38
Protanomalija	1,0	0,02
Deuteranopija	1,1	0,01
Protanopija	1,0	0,02
Tritanopija	0,002	0,001

3 Pregled obstoječih metod

Daltonizacija označuje postopke prebarvanja slik za barvno slepe. Vsebinsko neodvisne metode globalno prilagajajo piksele, so hitre in primerne za dikromate, vendar pogosto preveč spremenijo barve in niso ustrezne za anomalne trikromate.

Vsebinsko odvisne metode upoštevajo lokalno strukturo slike, zato dajejo boljše rezultate, a so računsko zahtevnejše. Kljub prilagoditvam ostaja izziv ohranjanja naravnega videza, zlasti pri blagih oblikah anomalne trikromatije.

Milić v (Milić, 2016) in Flatla v (Flatla, 2013) sta poskušala izboljšati vsebinsko odvisne daltonizacijske metode z vključitvijo testa barvne slepote in postopkom prebarvanja, ki ohranja naraven videz slike. Oba pristopa temeljita na treh korakih: segmentaciji slike v barvne centre, prepoznavanju konfuznih centrov in iskanju ustreznih premikov barv. Test Nade Milić določa konfuzne limite vzdolž glavnih linij konfuznosti v desetih diskretnih korakih, kar omogoča hitro, a manj natančno določitev globalne elipse. Segmentacija je izvedena z algoritmom K-povprečij, pri čemer upošteva le odtенок in nasičenost.

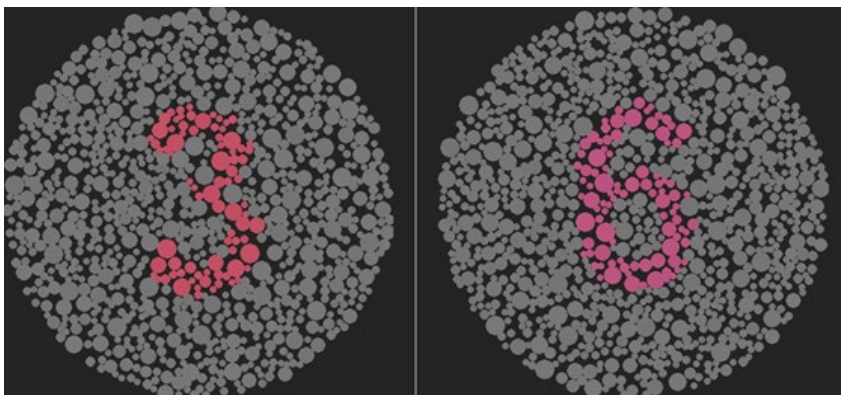
Prebarvanje barvnih centrov je odvisno od vrste barvne slepote. Pri dikromatiji in težjih oblikah anomalne trikromatije se uporabi tipska daltonizacija, kjer se centri prerazporedijo tako, da so čim bolj razmaknjeni, premiki pa potekajo pravokotno na linije konfuznosti. Pri blažjih oblikah se uporabi jakostna daltonizacija, kjer velikost premika določa globalna elipsa; centri na isti elipsi se premaknejo tako, da znotraj ostane le eden. Piksli se nato prebarvajo z ohranjanjem razmerij med barvami. Milić izpostavi, da veliko število centrov vodi v prekrivanje in slabše rezultate, svetlost pa v postopku ni upoštevana.

Flatla uporablja segmentacijo na podlagi pogostosti posameznih barv, pri čemer del najpogostejših postanejo centri, ki ne vključujejo podobnih odtenkov. To lahko privede do ustvarjanja preveč centrov in izključitve redkejših, a za uporabnika pomembnih barv. Limite barvne slepote določa z binarnim iskanjem po linijah konfuznosti in svetlosti, nato oblikuje globalni elipsoid in kopunkturno točko ter z njima prepozna konfuzne centre. Nadomestne barve poišče s simuliranim ohlajanjem, pri čemer optimizacija išče ravnotežje med ohranjanjem naravnosti in izboljšano razločnostjo. Nekonfuzne barve ostanejo nespremenjene, zato slika ostane naravna za normalnovidne in hkrati bolj razločna za barvno slepe.

4 Predlagana rešitev

4.1 Test barvne slepote z izokromatskimi tablicami

Na podlagi izokromatskih tablic (glej sliko 1) izvedemo test, s katerim določimo barve, ki so za uporabnika konfuzne, ne glede na vrsto barvne slepote ali vpliv okolja. Barvne krogce različnih odtenkov naključno razporedimo po šabloni, pri čemer variramo položaj, velikost in svetlost, da zaznavo omogoča izključno razlika v odtenku. Barve primerjamo v prostoru CIELUV, kjer se razlike dobro ujemajo z zaznavo. Med testom spreminjamo odtenke vzdolž linij konfuznosti in svetlobne osi ter z binarnim iskanjem določimo osem diskriminacijskih limit – točke, kjer barva preide iz neločljive v ločljivo glede na ozadje.



Slika 1: Izokromatske tablice.

Vir: lasten.

Iz izmerjenih limit zgradimo tridimenzionalni globalni elipsoid, kjer velika os predstavlja dominantno vrsto barvne slepote, manjša pa vpliv nedominantnih vrst in okolja. Zaradi uniformnosti barvnega prostora CIELUV lahko elipsoid posplošimo na vse barve v prostoru. Za primerjavo dveh barv namesto premikanja in rotiranja elipsoida transformiramo sekundarno barvo v lokalni koordinatni sistem elipsoida. Najprej izračunamo kot rotacije glede na kopunkturno točko po enačbi (1), kjer sta u^k in v^k koordinati kopunkturne točke, u^* in x^* pa koordinati primarne barve. Iz kota sestavimo rotacijsko matriko (enačba (1)) in z njo rotiramo sekundarno barvo. Nato uporabimo standardno enačbo elipsoida in preverimo, ali transformirana barva leži v njegovi notranjosti – v tem primeru sta barvi konfuzni.

$$\begin{aligned} \Theta &= \arctan(u^k - u^*, v^k - v^*) \\ R &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\Theta) & -\sin(\Theta) \\ 0 & \sin(\Theta) & \cos(\Theta) \end{bmatrix} \end{aligned} \quad (1)$$

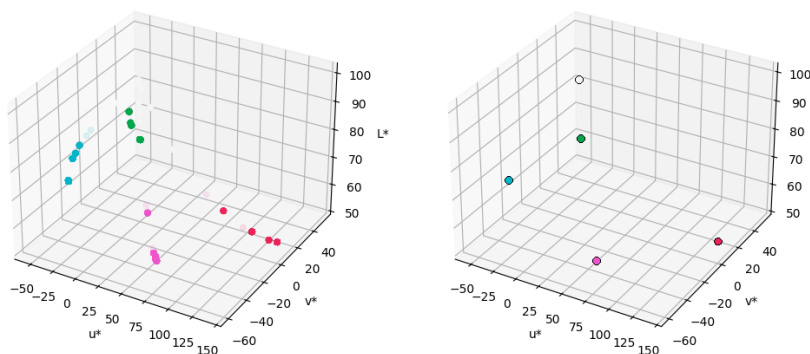
4.2 Prebarvanje slik

Postopek prebarvanja temelji na štirih korakih: i) izbiri reprezentativnih barv, ii) zaznavi konfuznih parov, iii) optimizaciji nadomestnih centrov in iv) rekonstrukciji slike. Ker slika vsebuje veliko barvnih vrednosti, najprej v prostoru CIELUV izvedemo segmentacijo z metodo Mean Shift. Ta iterativno premika točke proti lokalnim maksimumom gostote. Premik je določen z vektorjem iz enačbe (2), kjer $S_h(x)$ predstavlja množico točk v jedru širine h . Ko je premik zanemarljiv, točka doseže stabilno območje gostote, ki postane barvni center.

$$M_h(x) = m_h(x) - x \quad m_h(x) = \frac{1}{|S_h(x)|} \sum_{x_i \in S_h} x_i \quad (2)$$

Postopek pohitrimo z mrežnim grupiranjem barvnega prostora: prostor razdelimo na celice velikosti jedra in v vsaki hranimo le število točk ter vsoto njihovih koordinat, zato se premiki računajo nad celicami namesto nad vsemi piksli. Širino jedra ocenimo iz unikatnih kromatičnih centrov, začetna semena pa neodvisno iščejo lokalne maksimume gostote. Po konvergenci združimo centre, ki so bližje od polovice širine jedra, ter odstranimo majhne segmente. Piksle nato dodelimo najbližjemu centru s postopkom k-najbližjih sosedov (implementacija z drevesom K-D). Rezultat sta kompaktna množica barvnih centrov in pripadnosti pikslov posameznim centrom (glej sliko 2).

V naslednjem koraku med barvnimi centri poiščemo zaznavno nerazločljive pare. Za vsak center ostale centre preslikamo v njegov lokalni koordinatni sistem, orientiran glede na kopunkturno točko dominantne vrste barvne slepote, ter s standardno enačbo elipsoida preverimo, ali ležijo v njegovi notranjosti. Če da, sta barvi konfuzni. Tako dobimo graf zaznavne nerazločnosti, kjer povezave predstavljajo pare barv, ki jih uporabnik ne razlikuje.



Slika 2: Barvne točke slike (levo) in najdeni centri (desno), predstavljeni v barvnem prostoru CIELUV.

Vir: lasten.

Odstranjevanje konfuznosti oblikujemo kot optimizacijski problem premika barvnih centrov v prostoru CIELUV. Novi centri morajo ostati znotraj sRGB in hkrati zunaj elipsoidov vseh drugih centrov, pri čemer zahtevamo razdaljo, večjo od roba elipsoida. Optimizacijo izvedemo v dveh fazah: najprej premaknemo le centre z največ konflikti oziroma najmanjšim deležem pikslov, nato po potrebi vse konfuzne centre. Stroškovna funkcija minimizira vsoto kvadratov premikov, pri čemer je sprememba svetlosti L^* dodatno utežena. Problem rešujemo z metodo SLSQP, ki omogoča upoštevanje nelinearnih omejitev elipsoida in barvnega prostora.

V zadnjem koraku sliko rekonstruiramo tako, da piksli nespremenjenih centrov ohranijo izvirne barve, piksli premaknjenih centrov pa se prebarvajo glede na razdaljo od novega centra v prostoru CIELUV. Piksli blizu centra prevzamejo njegovo novo barvo, bolj oddaljeni pa se premaknejo v smeri vektorja med starim in novim položajem. Takšna interpolacija ohrani lokalne prehode in naraven videz, hkrati pa odstrani vse zaznavno nerazločljive barvne pare.

5 Rezultati in diskusija

Sorodne metode običajno ne poročajo časovne zahtevnosti, kar kaže, da niso zasnovane za realni čas. Flatla (Flatla, 2013) navaja povprečno 0,005 ms na primerjavo, izračunano nad milijonom parov barv RGB. Naša funkcija primerjave barv se uporablja pri iskanju konfuznih parov in znotraj optimizacije SLSQP, zato smo jo pospešili z zgodnjim preverjanjem razlike v svetlosti L^* . Če je ta večja od

krajše polosi elipsoida, nadaljnji izračuni niso potrebni. Meritve nad milijonom parov kažejo bistveno hitrejšo delovanje: na posamezno primerjavo v povprečju približno 0,000105 ms v SLSQP in 0,000314 ms pri samostojnem iskanju, kar pomeni skupno 105,42 ms oziroma 313,81 ms. Pohitritev izhaja iz vektorizacije, zgodnjega preverjanja svetlosti in dejstva, da transformiramo le sekundarno barvo, ne pa celotnega elipsoida.

Za oceno uporabnosti smo preverili vpliv velikosti slike na čas izvajanja. Preizkusili smo 50 slik (npr. naravne scene, zemljevide, Ishiharine tablice) v treh velikostih (400×400, 800×800, 1920×1080), vsako desetkrat. Rezultati (tabela 2) kažejo majhne standardne odklone in približno linearen porast časa glede na število pikslov, kar potrjuje, da na hitrost najbolj vpliva velikost slike. Tudi najdaljši časi ostajajo v območju, ki še omogoča dobro uporabniško izkušnjo.

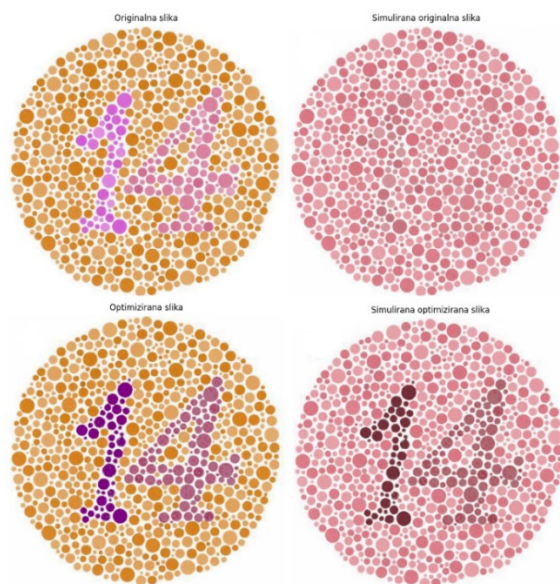
Tabela 2: Hitrost izvedbe postopka daltonizacije glede na velikost slike v pikslih

	400 x 400	800 x 800	1920 x 1080
Povprečni čas	1,15 s	3,72 s	12,53 s
Standardni odklon	0,22 s	0,41 s	1,35 s

Učinkovitost metode smo preverili na tablicah Richmond HRR, kjer barve vzorca in ozadja ležijo na isti liniji konfuznosti in po simulaciji postanejo skoraj enake. Po optimizaciji morajo biti številke razločne tako pred kot po simulaciji. Na testni tablici za tritanopijo (glej sliko 3) original prikazuje oranžno ozadje in škrlatno-rožnate krogce (1 in 4), ki po simulaciji niso več ločljivi. Naša optimizacija številke ponovno jasno razloči tudi po simulaciji, kar potrjuje odpravo konfuznosti. Primer je ilustrativen, saj tablice vsebujejo malo barv in ciljajo predvsem na dikromatijo.

Metodo smo primerjali z rešitvijo Milić (2019) na zemljevidu podzemne železnice s petimi progami (glej sliko 4). Za protanope sta konfuzni progi M1 (RGB: 155,155,25) in M2 (RGB: 85,165,30). Naša segmentacija je določila sedem centrov in zaznala konfuznost le med M1 in M2, zato smo spremenili le barvo M1 v RGB (131,167,100). Sorodna metoda je spremenila več centrov in večji del slike. Razlike smo kvantitativno ocenili s povprečno barvno razliko po enačbi (3), ki meri evklidsko razdaljo med originalnimi in prebarvanimi centri v prostoru CIELUV.

$$\Delta E_{L*u*v*} = \frac{1}{n} \sum_{i=1}^n \Delta E_{Luv}(c_{1,i}, c_{2,i}) \quad (3)$$



Slika 3: Simulacija tablice Richmond HRR za zaznavanje tritanopije.

Vir: lasten.

V naši metodi se je prebarval le center proge M1 z barvno razliko $\Delta E_{L*u*v*} = 31,91$, povprečje vseh petih prog pa znaša $\Delta E = 6,38$. V sorodni metodi, kjer so bile spremenjene vse proge, povprečje doseže $\Delta E = 45,30$. Čeprav so posamezne spremembe barv manj izrazite, njihovo število močno vpliva na celoten videz slike. Vizualno je razvidno, da naša metoda bistveno manj posega v izvirne barve, hkrati pa ohrani razločljivost prog za barvno slepe opazovalce.

Naravne slike (glej sliko 5) smo vključili zato, ker prav pri njih daltonizacijske metode najtežje ohranijo kakovost. Zaradi velikega števila odtenkov in zveznih prehodov lahko zmanjšanje števila barvnih centrov povzroči opazne preskoke in artefakte. Več centrov bi te težave zmanjšalo, vendar bi močno povečalo časovno zahtevnost zaznavanja konfuznih parov in optimizacije, zato gre za kompromis med kakovostjo in hitrostjo izvajanja.



Slika 4: Primerjava lastne in sorodne rešitve na sliki zemljevida.

Vir: lasten.

6 Sklep

Barvna slepota otežuje dostop do informacij, kadar so te podane z barvami, zato smo predstavili personaliziran pristop daltonizacije, ki temelji na individualnih diskriminacijskih mejah, izmerjenih z izokromatskimi tablicami. Iz teh podatkov v prostoru CIELUV zgradimo globalni elipsoid zaznavne podobnosti, ki omogoča natančno določanje konfuznih barvnih parov in upošteva različne oblike barvne slepote ter vplive okolja. Postopek nato vključuje izbor reprezentativnih barv, zaznavo konfuznih parov, optimizacijo nadomestnih barv z omejitvami prostora sRGB ter končno preslikavo na sliko. Cilj ni le povečanje kontrasta, temveč ohranitev naravnega videza ob hkratni izboljšani razločljivosti informacij.



Slika 5: Primer prebarvanja naravnih slik.

Vir: lasten.

Rezultati kažejo, da naša metoda ponuja bolj uravnoteženo prilagoditev kot splošne rešitve, saj hkrati izboljša razločevanje barv in omeji nepotrebne spremembe. Personaliziran pristop se tako izkaže kot primernejša smer razvoja sistemov za prilagajanje vizualnih vsebin, saj omogoča bolj dostopno in uporabniku prilagojeno okolje.

Viri in literatura

- Fairchild, M. D. (2005). *Color Appearance Models*, 2. izdaja, Chichester: John Wiley & Sons.
- Nathans, J. (1999). The Evolution and Physiology of Human Color Vision: Insights from Molecular Genetic Studies of Visual Pigments. *Nuron*, vol. 24, str. 299-312.
- Simunovic, M. (2009). Color vision deficiency. *Eye*, vol. 24, str. 747-755.
- Hunt, R. W. G. (2011). Pointer, M., *Measuring Color*, 4. izdaja, Chichester: John Wiley & Sons.
- Milić, N. (2016). Model optimizacije slike za korisnike sa poremećajima viđenja boja. Novi Sad: Univerzitet u Novom Sadu Fakultet tehničkih nauka grafičko inženjerstvo i dizajn.
- Flatla, D. R. (2013). Individualized models of color differentiation through situation-specific modelling. Saskatoon: Department of Computer Science University of Saskatchewan Saskatoon.

HUMAN ACTION RECOGNITION AND CUSTOM DATASET FOR BUS PASSENGER SAFETY

DONG GYU LEE, DONG SEOG HAN

Kyungpook National University, Electronic and Electrical Engineering,
Daegu, Republic of Korea
doe58@knu.ac.kr, dshan@knu.ac.kr

Although many buses are equipped with onboard CCTV to assist drivers in monitoring the cabin, identifying abnormal passenger conditions in real time remains challenging. This work aims to enable early detection of emergency situations by recognizing passenger behaviors from video captured by cameras installed inside special-purpose vehicles. In practice, the interior of bus-like vehicles is highly complex and cluttered, which makes robust behavior understanding difficult. Another major limitation is the lack of publicly available datasets that represent diverse special-vehicle interiors. To address these issues, we recreated multiple bus environments and collected in-vehicle data to build a dedicated dataset. Using the collected dataset, we benchmarked several deep learning models to explore suitable approaches for passenger behavior recognition, and our proposed method achieved higher recognition accuracy than the competing baselines.

DOI

[https://doi.org/
10.18690/um.feri.4.2026.10](https://doi.org/10.18690/um.feri.4.2026.10)

ISBN

978-961-299-111-1

Keywords:

human action recognition,
bus passenger safety,
deep learning,
custom dataset,
pose estimation



University of Maribor Press

DOI
[https://doi.org/
10.18690/um.feri.4.2026.10](https://doi.org/10.18690/um.feri.4.2026.10)

ISBN
978-961-299-111-1

Ključne besede:

prepoznavanje človeških
aktivnosti,
varnost potnikov na
avtobusu,
globoko učenje,
namenska podatkovna
zbirka,
ocena položaja telesa

PREPOZNAVANJE ČLOVEŠKIH AKTIVNOSTI IN RAZVOJ PODATKOVNE ZBIRKE ZA VARNOST POTNIKOV NA AVTOBUSU

DONG GYU LEE, DONG SEOG HAN

Nacionalna univerza Kyungpook, elektrotehnika in inženiring,
Deagu, Južna Koreja
doc58@knu.ac.kr, dshan@knu.ac.kr

Kljub temu da so številni avtobusi opremljeni z video-nadzornimi sistemi za pomoč voznikom pri spremljanju dogajanja v vozilu, pravočasna prepoznavna nepredvidenih dogodkov, ki ogrožajo varnost potnikov, še vedno predstavlja velik izziv. Cilj tega dela je zgodnje odkrivanje nevarnih situacij s prepoznavanjem vedenja potnikov na podlagi videoposnetkov iz kamer, nameščenih v vozilih. Zanesljivo prepoznavanje človeških aktivnosti otežuje kompleksna notranjost vozil, kjer oprema in drugi potniki povzročajo številne okluzije. Dodatno omejitev predstavlja pomanjkanje javno dostopnih podatkovnih zbirk, ki bi zajemale raznolike pogoje v vozilih. Za reševanje teh izzivov smo zasnovali več različnih avtobusnih okolij in zbrali posnetke iz notranjosti vozil, s katerimi smo oblikovali novo podatkovno zbirko. Na podlagi zbrane zbirke smo izvedli primerjavo več pristopov globokega učenja za prepoznavo človeških aktivnosti, pri čemer naša metoda dosega najvišjo točnost v primerjavi z obstoječimi metodami.



Univerzitetna založba
Univerze v Mariboru

1 Introduction

In large vehicles such as buses, passengers may be injured when they collide with surrounding structures or lose balance during sudden acceleration or abrupt braking, as well as when they move inside the cabin while the vehicle is in motion. The severity of these incidents ranges from minor bruises to serious falls, and the risk can be higher for elderly passengers or those with medical conditions who may experience unexpected physical distress. Although many buses are equipped with onboard CCTV and safety fixtures such as handrails and grab handles to mitigate these risks, ensuring passenger safety remains a persistent challenge.

This study focuses on understanding passenger behavior from bus-mounted CCTV footage. Specifically, we aim to detect passengers and infer their current actions so that potentially hazardous or abnormal situations can be communicated to the driver in a timely manner. To this end, we evaluate multiple deep learning-based approaches for passenger behavior recognition and compare their performance to identify the most suitable method.

A key difficulty is that real bus interiors are visually cluttered and frequently involve occlusions caused by seats, poles, and other passengers. Moreover, publicly available datasets that capture passenger behaviors under such complex in-bus conditions are scarce, making model training and fair evaluation difficult. To address this limitation, we constructed a bus-like environment and collected image data of staged passenger actions, which we then used for training and benchmarking the proposed recognition models.

1.1 Related Work

Human Action Recognition (HAR) has been studied extensively due to its reliance on spatiotemporal cues and its broad applicability in surveillance, robotics, sports analytics, healthcare, and human-machine interaction. As deep learning architectures diversified, recent surveys have focused on organizing the field and clarifying the interplay between model families, sensing modalities, and evaluation protocols.

SMART-Vision provides a vision-centric survey and introduces a Venn-diagram taxonomy that explicitly captures hybridization among major paradigms such as two-stream models, 3D CNNs, graph-based methods, motion-centric designs, and transformers (AlShami et al., 2025). In addition to general RGB-based HAR, a dedicated survey of RGB-D action recognition summarizes algorithms and benchmarks that leverage depth information for improved robustness under viewpoint and illumination variations (Zhang and Wang, 2025). Few-shot action recognition has also been reviewed systematically, categorizing methods (e.g., generative and meta-learning based approaches) and emphasizing video-specific challenges such as temporal modeling and high-dimensional representations (Wanyan et al., 2024). Together, these survey works motivate a structured discussion of architectures and generalization settings in modern HAR.

Although video is the dominant input for HAR, action understanding from still images or sparse key frames remains important when temporal evidence is limited. Transfer-learning pipelines that fine-tune ImageNet-pretrained CNN backbones provide practical baselines for still-image action classification, demonstrating that strong spatial representations can partially compensate for missing motion cues (Kanangama and Thirukumaran, 2024). For sports scenarios, CBAM-enhanced ResNet variants combined with keypoint detection have been explored to emphasize discriminative spatial regions and pose-related patterns in video frames (Xu and Sun, 2024). Beyond classification, image-based action retrieval has been formulated as a distinct task: region-aware transformer models fuse person-centric and contextual features to retrieve semantically similar actions from static imagery (Wang et al., 2024). These studies highlight the value of region/context modeling when temporal information is weak or absent.

For video-based HAR, a key challenge is capturing temporal dynamics efficiently while maintaining discriminative capacity. Temporal Shift Module (TSM)-based backbones offer a lightweight mechanism for temporal interaction, and an ensemble-driven TSM system achieved strong performance in a competitive multi-modal recognition setting (Duong and Gomez-Krämer, 2024). Recent work also emphasizes fine-grained action details. FocusVideo introduces learnable local action queries to attend to action-relevant regions without explicit region annotations and couples global context with local refinement in transformer blocks, improving recognition for detail-sensitive classes (Wang et al., 2025).

Skeleton-based methods exploit human body structure via graph modeling to represent joints and their interactions. Multi-scale spatial graph convolution has been proposed to better capture both local and long-range joint dependencies by operating across multiple receptive fields (Fang and Wang, 2025). Beyond skeleton-only pipelines, hybrid fusion approaches aim to combine complementary cues from raw pixels and structured pose graphs. GCCIN adopts a dual-path architecture that fuses CNN-derived pixel features with ST-GCN-style skeleton features for multi-scale action recognition, seeking improved robustness under complex appearance variations (Huang and Zhang, 2025).

Deployment-oriented HAR often faces challenging capture conditions and heterogeneous sensors. Drone-based action recognition pipelines address aerial viewpoints characterized by motion blur, dynamic backgrounds, and scale changes (Abbas and Jalal, 2024). For low-illumination settings, Dark Transformer studies cross-domain adaptation with video transformers to transfer knowledge from source domains to actions observed “in the dark” (Ulhaq, 2024). In distributed multi-camera environments with weak supervision, action selection learning has been proposed for multi-label multi-view recognition, learning to select action-relevant frames and to fuse wide-range views using only video-level tags (Nguyen et al., 2024). Finally, multimodal fusion of video and wearable IMU signals has been investigated in sports analytics via CNN–LSTM pipelines with adaptive weighting mechanisms, demonstrating that complementary sensing can improve recognition and enable quantitative motion assessment (Chen et al., 2025).

Overall, the above literature indicates that modern HAR systems are increasingly hybrid in architecture and modality, and that robustness and generalization are central considerations for real-world deployment.

2 Experiments

Due to the lack of publicly available passenger-safety datasets that reflect the characteristics of Korean bus interiors (city bus, express bus, emergency bus), we built a dedicated acquisition environment and collected our own data. Four cameras were mounted inside a test bus to capture passenger behaviors from multiple viewpoints. Five participants acted according to predefined scenarios designed to cover both a single-passenger setting and a crowded setting with multiple passengers, enabling us to model behavior under different levels of interaction and occlusion.

During recording, each participant repeatedly performed the target actions, and fixed time intervals were introduced between actions to ensure clear temporal separation for labeling.

To reflect realistic in-bus conditions, we intentionally included segments where the full body of a passenger is clearly visible as well as segments where the subject is partially occluded by seats, handrails, or other people. We defined five behavior classes commonly observed in buses: sitting, standing, holding (grabbing), walking, and falling. After data collection, we developed an in-house preprocessing tool to convert the raw videos into a format suitable for our recognition pipeline. For each camera, recordings were captured from diverse angles with durations ranging from approximately 3 to 5 minutes, and stored as 45-frame clips. In total, the dataset contains about 14 GB of video data.



Figure 1: Dataset for various bus environments

Source: Own.

All experimental models were trained and tested using our collected custom dataset. For a fair comparison, we resized the input to 256×192 pixels and used approximately 16K training images in total. Performance was evaluated in terms of Average Precision (AP).

Among the baselines, DeepPose (Toshev and Szegedy, 2014) showed the weakest results, largely because it relies on a relatively simple CNN design; its accuracy dropped substantially when the target person was partially occluded by surrounding

objects. HRNet (Wang and Sun K,2020), which preserves high-resolution representations through parallel high/low-resolution branches and repeated feature exchange, did not achieve the expected gains in our setting. We attribute this to the constrained input resolution, which likely limited HRNet’s ability to exploit fine-grained spatial cues.

We also evaluated SimCC (Li Yang et al, 2022), which estimates joints via coordinate classification rather than heatmap regression. Since SimCC requires an additional detector, we adopted an HRNet-based detector in our implementation. Representing joint locations using bin-style coordinates yielded higher performance than vanilla HRNet, presumably because it mitigates quantization errors commonly introduced by heatmap discretization.

Table 1: Performance comparison table

	Input size	AP	AP^{50}	AP^{75}
DeepPose	256 x 192	0.499	0.664	0.523
HRNet	256 x 192	0.540	0.677	0.631
SimCC	256 x 192	0.671	0.799	0.736
YOLO-Pose	256 x 192	0.704	0.849	0.748
RTMPose	256 x 192	0.721	0.866	0.795
PS-Pose(ours)	256 x 192	0.725	0.877	0.802

YOLO-Pose (Maji and Nagori,2022), slightly outperformed SimCC. This improvement appears to stem from differences in both detector quality and the training strategy described in the original method. Moreover, by leveraging CSP-Darknet and PANet with multi-head predictions, YOLO-Pose remained relatively robust under occlusion-heavy scenes.

RTMPose (Jiang et al., 2025), incorporating more recent design choices, achieved the strongest overall performance in our experiments and was particularly suitable for crowded bus scenarios with frequent occlusions. Interestingly, YOLO-Pose produced better AP^{50} scores, which we interpret as an effect of detector disparity—our RTMPose configuration used YOLOv3 as the detector. Despite this, RTMPose provides a practical balance by combining advantages of coordinate classification (computational efficiency) with strong pose estimation accuracy, making it scalable for real-world deployment.

Finally, we considered PS-Pose, an enhanced variant built upon RTMPose. In our study, we integrated an additional module designed to exploit surrounding contextual cues in crowded environments, where interactions with nearby objects and passengers are common. This context-aware design improved discrimination between visually similar behaviors, leading to a measurable gain in recognition accuracy under multi-passenger and occlusion-dominant conditions.

3 Conclusion

Using bus-mounted cameras, this study collected scenario-driven images under diverse conditions-including single-passenger and crowded-cabin cases-to build a dataset tailored to complex in-bus environments. We then investigated effective approaches for recognizing passenger behaviors by benchmarking a skeleton-based keypoint method against a spatial-information fusion model and comparing their performance.

References

- Abbas, Y., & Jalal, A. (2024). Drone-based human action recognition for surveillance: a multi-feature approach. In 2024 International Conference on Engineering & Computing Technologies (ICECT)(pp. 1–6). doi:10.1109/ICECT61618.2024.10581378
- AlShami, A. K., Rabinowitz, R., Lam, K., Shleibik, Y., Mersha, M., Boulton, T., & Kalita, J. (2025). Smart-vision: survey of modern action recognition techniques in vision. *Multimedia tools and applications*, 84(27), 32705–32776. doi:10.1007/s11042-024-20484-5
- Benavent-Lledo, M., Mulero-Pérez, D., Ortiz-Perez, D., Garcia-Rodriguez, J., & Argyros, A. (2024). Enhancing action recognition by leveraging the hierarchical structure of actions and textual context. *arXiv preprint arXiv:2410.21275*. <https://arxiv.org/abs/2410.21275>
- Chen, Y., Chen, Y., Lou, K., Chen, Y., Li, H., & Jiang, Z. (2025). Multimodal Action Recognition and Quantitative Analysis Model for Pickleball Based on CNN-LSTM with Adaptive Weight Fusion. In 2025 4th International Conference on Image Processing, Computer Vision and Machine Learning (ICICML)(pp. 855-866). doi:10.1109/ICICML67980.2025.11333691
- Duong, A. K., & Gomez-Krämer, P. (2024). Action recognition using temporal shift module and ensemble learning. In *International Conference on Pattern Recognition*(pp. 302–313). doi:10.1007/978-3-031-88217-3_23
- Fang, Y., & Wang, B. (2025). Multi-scale Spatial Graph Convolutional Network for Skeleton-based Action Recognition. In 2025 2nd International Conference on Digital Image Processing and Computer Applications (DIPCA)(pp. 247-250). doi:10.1109/DIPCA65051.2025.11042736
- Huang, W., & Zhang, J. (2025). Graph Convolution Combined with Image Pixel Information Networks for Multi-Scale Action Recognition. In 2025 5th International Conference on Artificial Intelligence, Automation and High Performance Computing (AIAHPC)(pp. 157-163). doi:10.1109/AIAHPC66801.2025.11290605
- Jiang, T., Lu, P., Zhang, L., Ma, N., Han, R., Lyu, C., Li, Y., & Chen, K. (2023). RTMPose: Real-Time Multi-Person Pose Estimation based on MMPose. *arXiv preprint arXiv:2303.07399*. doi:10.48550/arXiv.2303.07399.

- Kanangama, K. A. N. C., & Thirukumaran, S. (2024). Enhancement of Image Based Human Action Recognition using Transfer Learning with Pre-Trained CNN Architectures. In 2024 9th International Conference on Information Technology Research (ICITR)(pp. 1-6). doi:10.1109/ICITR64794.2024.10857746
- Lan, J., & Yuan, T. (2024). V-APT: Action Recognition Based on Image to Video Prompt-Tuning. In 2024 2nd International Conference on Mechatronics, IoT and Industrial Informatics (ICMII)(pp. 63-67). doi:10.1109/ICMII62623.2024.00018
- Li, S., Li, W., Ren, Q., & Lin, D. (2025). Few-Shot Human Action Recognition with CLIP-Based Semantic-Guided Network. In 2025 5th International Conference on Computer, Control and Robotics (ICCCR)(pp. 573-577). doi:10.1109/ICCCR65461.2025.11072638
- Li, Y., Yang, S., Liu, P., Zhang, S., Wang, Y., Wang, Z., Yang, W., & Xia, S.-T. (2022). SimCC: A Simple Coordinate Classification Perspective for Human Pose Estimation. *Computer Vision – ECCV 2022*. doi:10.48550/arXiv.2107.03332. (Available at arXiv:2107.03332).
- Maji, D., Nagori, S., Mathew, M., & Poddar, D. (2022). YOLO-Pose: Enhancing YOLO for Multi Person Pose Estimation Using Object Keypoint Similarity Loss. *arXiv preprint arXiv:2204.06806*. doi:10.48550/arXiv.2204.06806.
- Nguyen, T. T., Kawanishi, Y., Komamizu, T., & Ide, I. (2024). Action Selection Learning for Multi-label Multi-view Action Recognition. In *Proceedings of the 6th ACM International Conference on Multimedia in Asia*(pp. 1–1). doi:10.1145/3696409.3700211
- Toshev, A., & Szegedy, C. (2014). DeepPose: Human Pose Estimation via Deep Neural Networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. doi:10.1109/CVPR.2014.214. (Available at arXiv:1312.4659).
- Ulhaq, A. (2024). Dark Transformer: A Video Transformer for Action Recognition in the Dark. *arXiv preprint arXiv:2407.12805*. <https://arxiv.org/abs/2407.12805>
- Wang, H., Zhao, J., & Gui, J. (2024). Region-aware image-based human action retrieval with transformers. *Computer Vision and Image Understanding*, 249, 104202. doi:10.1016/j.cviu.2024.104202
- Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., et al.(2020). Deep High-Resolution Representation Learning for Visual Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*.doi:10.1109/TPAMI.2020.2983686. (Available at arXiv:1908.07919).
- Wang, M., Huang, Z., Kong, X., Shen, G., Dai, G., Wang, J., & Liu, Y. (2025). Action Detail Matters: Refining Video Recognition with Local Action Queries. In *Proceedings of the Computer Vision and Pattern Recognition Conference*(pp. 19132–19142). doi:10.1109/CVPR52734.2025.01782
- Wanyan, Y., Yang, X., Dong, W., & Xu, C. (2024). A comprehensive review of few-shot action recognition. *arXiv preprint arXiv:2407.14744*. <https://arxiv.org/abs/2407.14744>
- Xu, C., & Sun, L. (2024). Convolutional Block Attention Mechanism with ResNet18 for Sports Action Recognition based on Video Image Key Point Detection. In 2024 International Conference on Distributed Systems, Computer Networks and Cybersecurity (ICDSCNC)(pp. 1-5). doi:10.1109/ICDSCNC62492.2024.10941504
- Zhang, Y., & Wang, Y. (2025). A comprehensive survey on RGB-D-based human action recognition: algorithms, datasets, and popular applications. *EURASIP Journal on Image and Video Processing*, 2025(1), 15. doi:10.1186/s13640-025-00677-0

ZAZNAVA IN KLASIFIKACIJA BREZPILOTNIH LETALNIKOV S KAMERO IN RADARJEM

PANE ANDOV, PRIMOŽ SMOGAVEC, DUŠAN GLEICH

Univerza v Mariboru, Fakulteta za elektrotehniko, računalništvo in informatiko,
Maribor, Slovenija
pane.andov@um.si, primoz.smogavec@um.si, dusan.gleich@um.si

V prispevku je predstavljen realnočasovni sistem za zaznavo brezpilotnih letalnikov, ki združuje vizualno detekcijo na osnovi globokega učenja in napredno radarsko obdelavo Range–Doppler signalov FMCW. Vizualni del temelji na modelu YOLOv8, implementiranem na platformi NVIDIA Jetson Orin Nano z uporabo industrijske kamere Basler. Radarski podsistem, izveden na Raspberry Pi, vključuje Gaussovo glajenje, odštevanje ozadja ter detekcijo ciljev z uporabo pragovne metode in gručenja DBSCAN v prostoru razdalje. Med podsistemoma je vzpostavljena komunikacija UDP, ki omogoča časovno usklajevanje zaznav in njihovo združevanje znotraj definirane časovnega intervala. Eksperimentalni rezultati potrjujejo, da kombinirani pristop izboljša zanesljivost zaznave v primerjavi s posameznimi metodami, zlasti v zahtevnih okoljskih pogojih. Prispevek analizira arhitekturo sistema, zmogljivost izvajanja na robnih napravah ter omejitve in možnosti nadaljnjih izboljšav.

DOI
[https://doi.org/
10.18690/um.feri.4.2026.11](https://doi.org/10.18690/um.feri.4.2026.11)

ISBN
978-961-299-111-1

Ključne besede:
dron,
detekcija dronov,
radar,
Dopplerjev radar,
YOLOv8



Univerzitetna založba
Univerze v Mariboru

DOI
[https://doi.org/
10.18690/um.feri.4.2026.11](https://doi.org/10.18690/um.feri.4.2026.11)

ISBN
978-961-299-111-1

Keywords:

drone,
drone detection,
radar,
Doppler radar,
YOLOv8

DETECTION AND CLASSIFICATION OF UNMANNED AERIAL VEHICLES WITH CAMERA AND RADAR

PANE ANDOV, PRIMOŽ SMOGAVEC, DUŠAN GLEICH

University of Maribor, Faculty of Electrical Engineering and Computer Science,
Maribor, Slovenia
pane.andov@um.si, primoz.smogavec@um.si, dusan.gleich@um.si

This paper presents a real-time system for unmanned aerial vehicle (UAV) detection that combines vision-based deep learning with advanced FMCW Range–Doppler radar processing. The visual subsystem is based on the YOLOv8 model deployed on the NVIDIA Jetson Orin Nano platform using an industrial Basler camera. The radar subsystem, implemented on a Raspberry Pi, performs Gaussian smoothing, background subtraction, and target detection using threshold-based segmentation followed by DBSCAN clustering in the range domain. UDP communication between the subsystems enables temporal synchronization of detections and their fusion within a predefined time window. Experimental results demonstrate that the combined approach improves detection reliability compared to individual sensing modalities, particularly under challenging environmental conditions. The paper analyzes the system architecture, computational performance on edge devices, and discusses current limitations and potential directions for further improvements.



University of Maribor Press

1 Uvod

V zadnjih letih se je uporaba brezpilotnih letalnikov močno povečala na področjih, kot so logistika, nadzor infrastrukture, kmetijstvo in rekreativna raba. Hkrati pa se pojavljajo tudi varnostni izzivi, povezani z nedovoljeno uporabo dronov v bližini kritične infrastrukture, letališč ali javnih prireditev (Obaid idr., 2025). Zaradi majhnih dimenzij, nizke radarske odbojnosti in pogosto nepredvidljivega gibanja predstavlja zanesljiva detekcija takšnih ciljev tehnični izziv.

Vizualni sistemi na osnovi globokega učenja, kot so sodobni enostopenjski detektorji objektov, kot je YOLO (angl. *You Only Look Once*), omogočajo učinkovito zaznavo objektov v realnem času. Vendar je njihova zanesljivost močno odvisna od svetlobnih pogojev, vremenskih vplivov in oviranosti vidnega polja. Po drugi strani radarski sistemi, zlasti Dopplerjevi radarji, omogočajo zaznavo ciljev neodvisno od osvetlitve ter zagotavljajo neposredno informacijo o razdalji in radialni hitrosti. Kljub temu pa so radarske meritve lahko obremenjene s šumom, motnjami okolja in stacionarnim ozadjem.

Zaradi komplementarnih lastnosti obeh pristopov se v zadnjem času vse bolj uveljavlja večmodalna fuzija podatkov, ki združuje vizualne in radarske informacije z namenom izboljšanja robustnosti zaznave.

V tem prispevku predstavljamo realnočasovni večmodalni sistem za detekcijo brezpilotnih letalnikov, ki združuje vizualno detekcijo na osnovi modela YOLOv8 z napredno Range–Dopplerjevo radarsko obdelavo FMCW (Zhou idr., 2019). Radarski podsistem vključuje Gaussovo glajenje, odštevanje ozadja ter detekcijo ciljev z uporabo pragovne metode in gručenja DBSCAN (angl. *Density-Based Spatial Clustering of Applications with Noise*) (Ester idr., 1996). Vizualni in radarski podsistem sta povezana prek komunikacije UDP (angl. *User Datagram Protocol*), ki omogoča časovno usklajevanje zaznav in njihovo združevanje znotraj definirane časovnega intervala.

2 Teoretični pregled

2.1 Algoritem YOLO

Za vizualno detekcijo je bil uporabljen algoritem YOLO (angl. *You Only Look Once*) (Redmon idr. 2016), ki spada med enostopenjske detektorje objektov. To pomeni, da model v enem prehodu skozi nevronske mreže napove razred objekta, položaj in velikost omejitvenega okvirja, kar omogoča hitro izvajanje v realnem času.

Arhitektura algoritma YOLO temelji na modularni zasnovi, ki je razdeljena na tri glavne sklope: osnovno omrežje (angl. *Backbone*), povezovalni del (angl. *Neck*) in izhodni del (angl. *Head*).

Osnovno omrežje predstavlja globoka konvolucijska mreža, katere naloga je iz vhodne slike izvleči značilnice.

Povezovalni del združuje značilnice iz različnih nivojev. Izhodni del je zaključni del modela, ki izvaja končno napoved. Ima tri izhode za zaznavanje objektov različnih velikosti: en izhod za majhne objekte, enega za srednje velike objekte in enega za velike objekte.

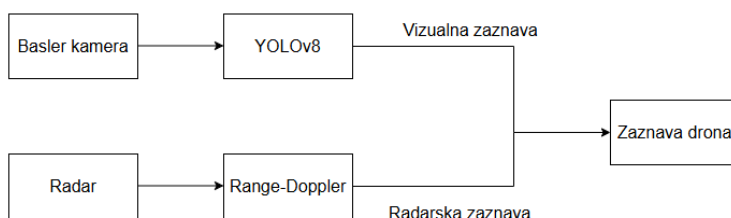
2.2 Radar

Radar deluje na principu oddajanja elektromagnetnega signala ter analize odbitega vala. Pri zaznavanju premikajočih se objektov ima ključno vlogo Dopplerjev pojav, pri katerem se frekvenca odbitega signala spremeni glede na relativno hitrost tarče.

V obravnavanem sistemu je uporabljen radar s frekvenčno moduliranim zveznim signalom (angl. *Frequency Modulated Continuous Wave*), ki oddaja linearno frekvenčno modulirane čirike (Eiadkaew idr., 2022). Razlika med oddanim in prejetim signalom omogoča določitev razdalje do cilja, medtem ko analiza spremembe faze med zaporednimi čiriki omogoča oceno radialne hitrosti. Z uporabo dvodimenzionalne Fourierjeve transformacije se izračuna Range-Dopplerjev spekter, ki prikazuje porazdelitev odbojev glede na razdaljo in hitrost.

3 Zasnova in izvedba sistema

Sistem, predstavljen v tej nalogi (glej sliko 1), je namenjen zaznavi brezpilotnih zrakoplovov s kombinacijo strojnega vida in Dopplerjevega radarja (Jian idr., 2018). Celotni sistem deluje kot kombinacija dveh neodvisnih vej, kjer je prva namenjena vizualni zaznavi brezpilotnih zrakoplovov z uporabo algoritma YOLO, druga pa predstavlja radarski sistem, ki temelji na Dopplerjevem pojavu za zaznavo premikajočih se objektov. Na koncu se rezultati obeh vej združijo in primerjajo. Če se zaznave iz obeh sistemov časovno ujemajo, zaznavo brezpilotnega zrakoplova štejejo za uspešno.



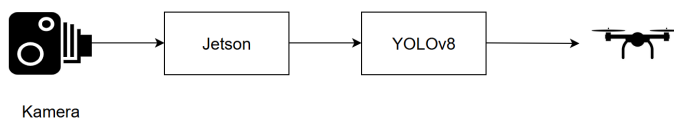
Slika 1: Zgradba celotnega sistema

Vir: lasten

3.1 Vizualni podsistem

Za optično zaznavo je bil uporabljen algoritem YOLOv8, ki spada med enostopenjske detektorje. To pomeni, da algoritem v enem samem koraku napove razred objekta, njegovo lokacijo in velikost. Tak pristop omogoča hitro obdelavo in je primeren za delovanje v realnem času.

Ena izmed prednosti YOLOv8 je, da se dobro obnese tudi pri zaznavanju majhnih in hitro premikajočih se objektov, kar je ključno pri detekciji dronov.



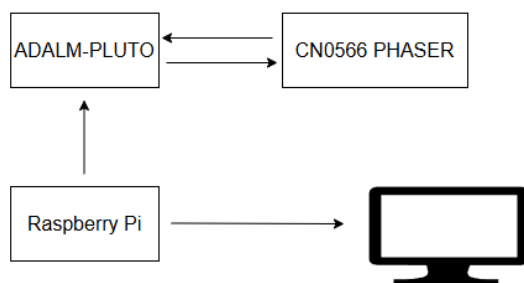
Slika 2: Zgradba vizualnega podsistema

Vir: lasten

Za zajem podatkov (slika 2) je bila uporabljena industrijska kamera Basler ace, ki omogoča stabilen zajem slik v realnem času in je primerna za robustne testne pogoje.

3.2 Radarski podsistem

Za zaznavo je bil uporabljen Dopplerjev radar FMCW, ki na podlagi frekvenčnega zamika odbitega signala določa razdaljo in radialno hitrost ciljev. Sistem sestavljajo ADALM-PLUTO SDR, razvojna plošča CN0566 in mikroračunalnik Raspberry Pi 4. ADALM-PLUTO deluje kot oddajno-sprejemni vmesnik RF (angl. *Radio Frequency*), CN0566 generira frekvenčno modulirane čirike, Raspberry Pi pa izvaja izračun Range-Dopplerjevega spektra z uporabo 2D Fourierjeve transformacije. Sistem deluje v realnem času ter omogoča konfiguracijo radarja, zajem in obdelavo podatkov (glej sliko 3).



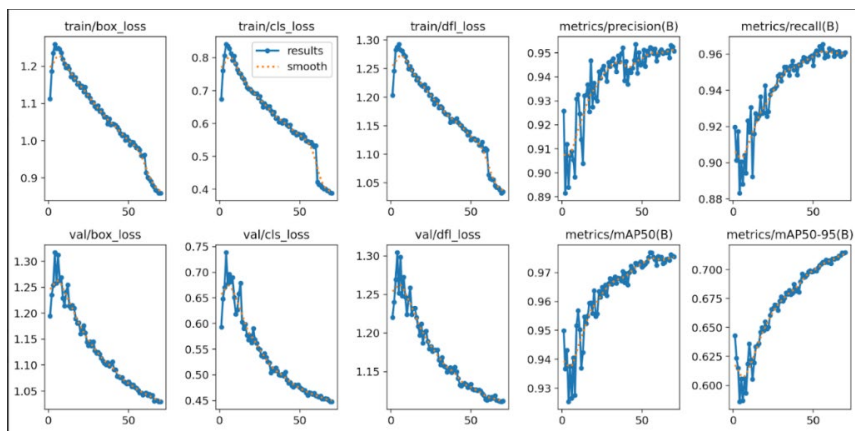
Slika 3: Zgradba radarskega podsistema
Vir: lasten

4 Eksperimentalni rezultati

4.1 Detekcija drona s kamero

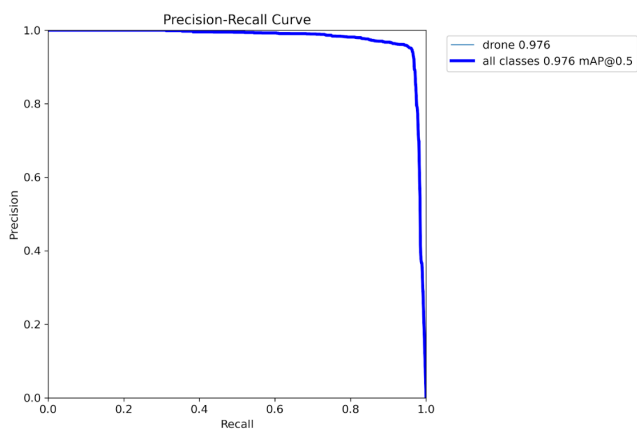
Za vizualno detekcijo drona je bil uporabljen model YOLOv8n (nano), ki je bil treniran na podatkovnem naboru z več kot 9000 označenimi slikami dronov. Učenje je potekalo 70 epoh pri vhodni ločljivosti 640 pik (slika 4).

Na krivuljah izgub vidimo, da te kažejo stabilno konvergenco brez izrazitega preučanja (angl. *Overfitting*). Končni rezultati so dosegli približno 95 % natančnost ter mAP nad 97 %, kar je več kot zadostno za realnočasovno uporabo (glej slike 5 in 6).



Slika 4: Rezultati učenja modela

Vir: lasten



Slika 5: Krivulja PR

Vir: lasten



Slika 6: Primer detekcije z algoritmom YOLO

Vir: lasten

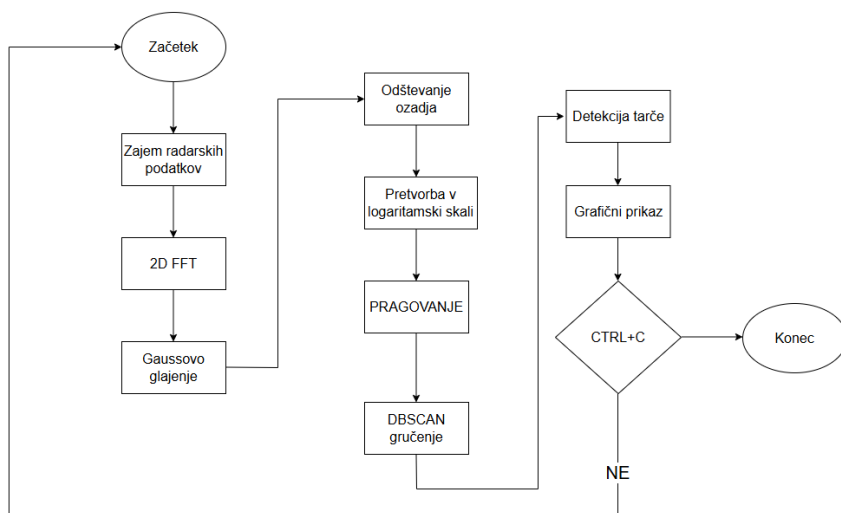
4.2 Detekcija drona z radarjem

Radar je zanesljivo zaznal premikajoči se dron, ki se je v spektru RD pojavil kot izrazit lokalni maksimum pri določeni razdalji in radialni hitrosti. Prisotni so bili tudi odboji statičnih objektov, zato je bil za zmanjšanje njihovega vpliva kot referenčno ozadje uporabljen prvi časovni okvir, zajet brez cilja, ki je bil odštet od vseh nadaljnjih meritev.

Na logaritmčno skalirani spekter RD je bil uporabljen prag, pri čemer so bile kot potencialne detekcije obravnavane točke nad izbranim pragom. Dodatne fizikalne omejitve (minimalna hitrost in omejeno območje razdalje) so zmanjšale število lažnih alarmov (glej sliko 7).

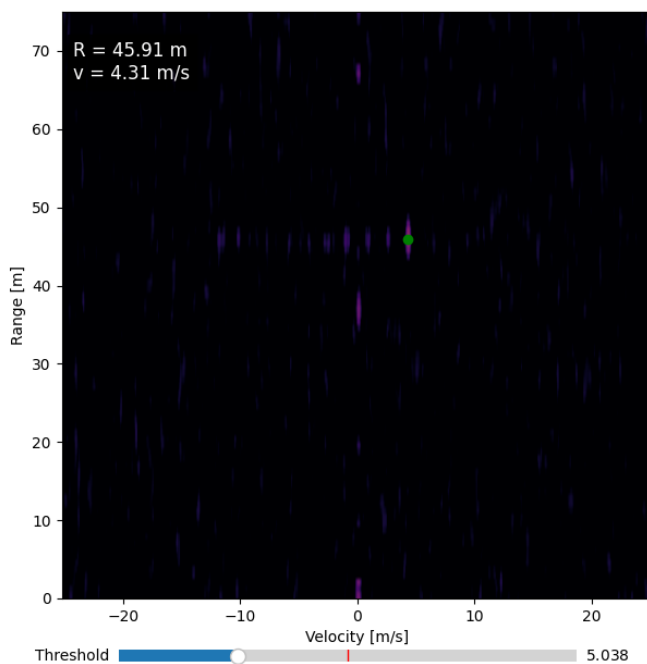
Ker se tarča v prostoru RD pojavi kot skupina energijskih elementov, je bil uporabljen algoritem DBSCAN za gostotno gručenje zaznanih točk. Za vsak grozd je bil izračunan uteženi centroid kot ocena razdalje in hitrosti cilja, pri čemer je bil kot primarna detekcija izbran grozd z največjo energijo (slika 8).

Rezultati potrjujejo robustnost radarskega podsistema pri zaznavanju brezpilotnega letalnika neodvisno od svetlobnih pogojev.



Slika 7: Zgradba vizualnega podsistema

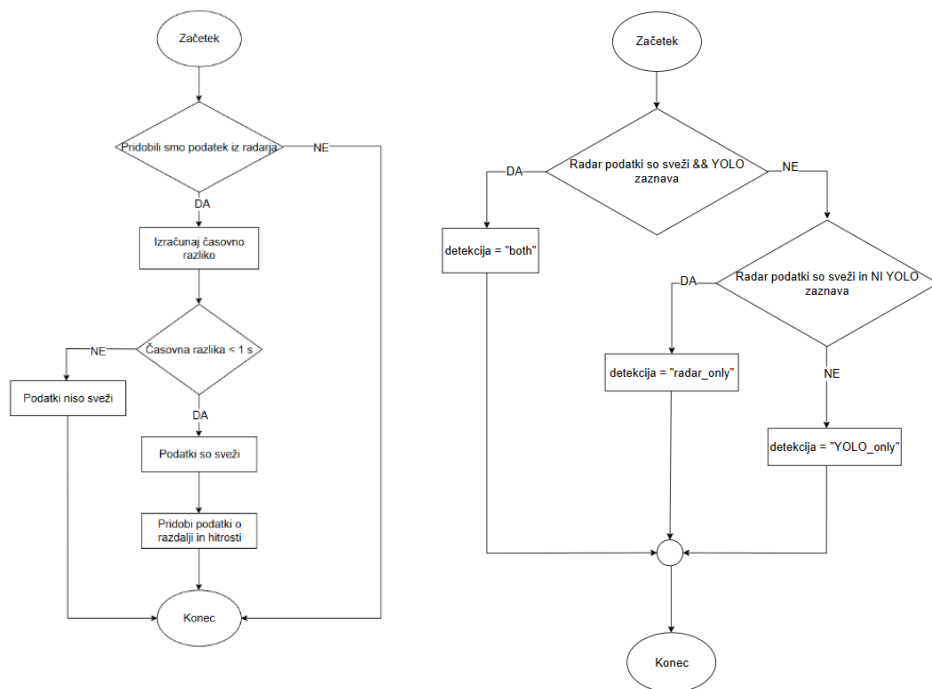
Vir: lasten



Slika 8: Primer detekcije v diagramu Range-Doppler
Vir: lasten

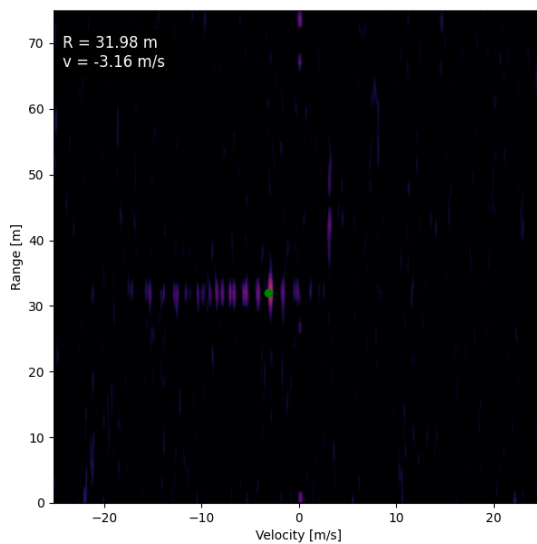
4.3 Časovna sinhronizacija obeh sistemov

Radar in vizualni sistem sta bila časovno usklajena z uporabo časovnih žigov (slika 9). Če je bila razlika med zaznavama manjša od 1 s, je bila zaznava klasificirana kot kombinirana. Sistem je razlikoval med tremi primeri: samo YOLO, samo radar ali kombinirana zaznava (glej sliko 10).



Slika 9: Diagrama poteka sinhronizacije radarske in optične sinhronizacije

Vir: lasten





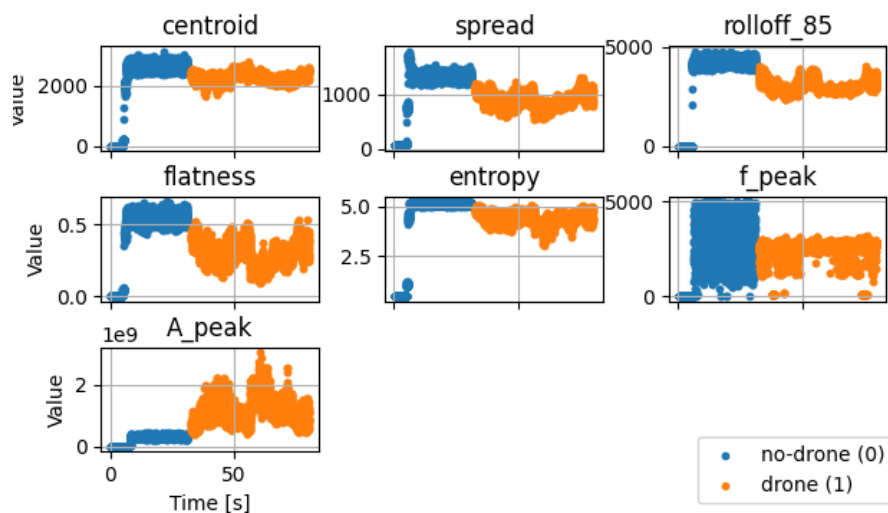
Slika 10: Sočasna detekcija z optičnim in radarskim sistemom
Vir: lasten

4.4 Klasifikacija na podlagi mikro-Dopplerjevih odtisov

Prisotnost drona se v sprejetem signalu odraža kot sprememba spektralne vsebine zaradi Dopplerjevega premika in mikro-Dopplerjevih pojavov, ki nastanejo zaradi rotacije propelerjev ter vibracij in gibanja strukture drona. Ti pojavi ustvarjajo značilne frekvenčne podpise, ki jih je mogoče uporabiti za detekcijo ali klasifikacijo.

V tem delu predstavimo sistem za zaznavanje prisotnosti drona na podlagi radarsko zajetih podatkov. Problem obravnavamo kot binarno klasifikacijo, kjer sistem odloča med dvema razredoma: prisotnost drona in odsotnost drona. Predstavlja osnovni korak, ki omogoča validacijo metodologije in bo v nadaljnjem delu raziskave razširjen na večrazredno klasifikacijo med več različnimi tipi dronov. Signal obdelamo s kratkočasovno Fourierjevo transformacijo (STFT), nato pa izvedemo klasifikacijo z dvema pristopoma z uporabo klasifikatorja SVM in 1D konvolucijske nevronske mreže (CNN), ki se uči neposredno iz spektralnih odtisih.

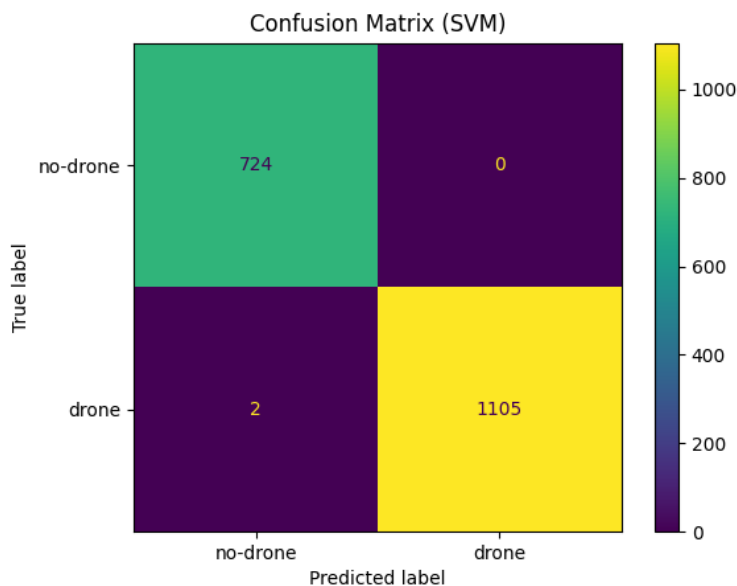
Prvi pristop temelji na ročno definiranih značilkah, izračunanih iz magnitude pozitivnega spektra STFT okvirja, in metodi podpornih vektorjev (SVM). Da zmanjšamo dimenzionalnost in se osredotočimo na relevantne komponente, značilke računamo le v območju 0–5000 Hz. V tem področju izberemo naslednje značilke: spektralno težišče, spektralno razpršenost, mejo spektralne energije (angl. roll-off, 85%), spektralno ravnost, spektralno entropijo, dominantno frekvenco in amplitudo vrha pri tej frekvenci. Spektralne značilke so prikazane na sliki 11.



Slika 11: Izbrane spektralne značilke

Vir: lasten

Podatki so razdeljeni na učno in testno množico s stratificirano delitvijo (70/30), da sta v obeh množicah enakomerno zastopana oba razreda (glej sliko 12).



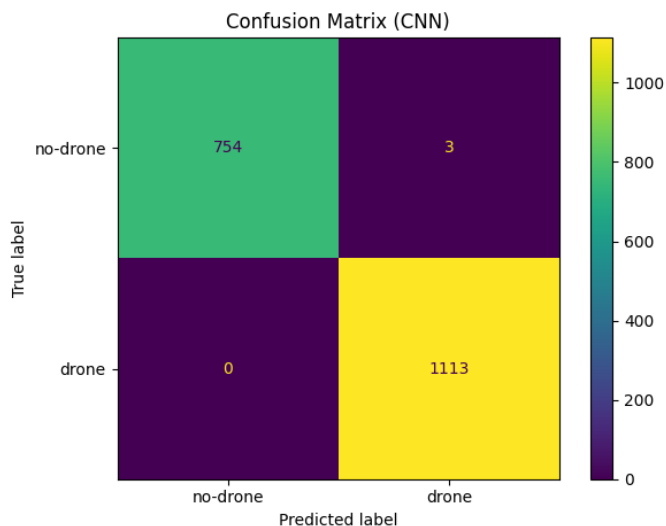
Slika 12: Matrika zmede pri učenju modela SVM.

Vir: lasten

Drugi pristop uporablja konvolucijsko nevronske mrežo, ki se uči neposredno iz spektra. Vhodni vektor sestavljajo magnitudni spektralni bini v območju 0–5000 Hz. Vhod CNN je oblike (1, F), kjer je F število frekvenčnih binov.

Uporabljena je enodimenzionalna konvolucijska nevronska mreža, sestavljena iz treh zaporednih konvolucijskih blokov ter zaključne polno povezane plasti. Prvi blok vključuje plast Conv1d (1→16, k=9), aktivacijsko funkcijo ReLU in operacijo MaxPool(2). Drugi blok je sestavljen iz Conv1d (16→32, k=9), ReLU in MaxPool(2). Tretji blok vsebuje Conv1d (32→64, k=9), ReLU ter prilagodljivo povprečno združevanje AdaptiveAvgPool(1). Končna reprezentacija se nato splošči (angl. Flatten) in posreduje v polno povezano plast Linear (64→1) za binarno klasifikacijo.

Mreža vrača napovedi (angl. logit) z, verjetnost dobimo s funkcijo sigmoid. Model učimo z optimizatorjem Adam ($\text{lr} = 1\text{e-}3$) ter funkcijo izgube BCEWithLogitsLoss. Učenje poteka 20 epoh.



Slika 13: Matrika zmede pri učenju modela CNN.

Vir: lasten

Matrika zmede za model CNN kaže zelo visoko uspešnost binarne klasifikacije. Model pravilno klasificira 754 primerov razreda 'brez drona' in 1113 primerov razreda 'dron' (slika 13). Zabeleženi so bili 3 lažno pozitivni primeri, medtem ko

lažno negativnih primerov ni bilo. Rezultati kažejo, da model CNN zelo učinkovito prepozna mikro-Dopplerjeve značilnosti drona in hkrati ohranja nizko stopnjo lažnih alarmov. Posebej pomembna je odsotnost lažnih negativnih primerov, saj ti v varnostnih aplikacijah predstavljajo kritično napako. Ta osnovna modela bosta služila kot osnova za večrazredno klasifikacijo med različnimi tipi dronov. Oba pristopa sta primerna za implementacijo v realnočasovni slapovni prikaz.

Takšna skoraj popolna ločitev med dvema razredoma potrjuje, da spektralni podatki vsebujejo dovolj diskriminativnih informacij za zanesljivo detekcijo prisotnosti drona. Vendar gre v tem primeru za binarni klasifikacijski problem. Pri razširitvi sistema na večrazredno klasifikacijo med različnimi tipi dronov ali drugimi zračnimi objekti bi se kompleksnost problema povečala, kar bi lahko vodilo do rahlega znižanja skupne točnosti ter več napačnih klasifikacij med medsebojno podobnimi razredi. Kljub temu trenutni rezultati predstavljajo trdno osnovo za nadaljnji razvoj večrazrednega modela, saj kažejo, da CNN in SVM uspešno zajameta ključne mikro-Dopplerjeve vzorce, na katerih je mogoče graditi naprednejše razločevanje.

7 Zaključek

V prispevku smo predstavili realnočasovni sistem za detekcijo brezpilotnih letalnikov, ki združuje vizualno zaznavo na osnovi globokega učenja in radarsko Range-Dopplerjevo obdelavo. Vizualni podsistem temelji na modelu YOLOv8, ki omogoča učinkovito detekcijo dronov v realnem času, medtem ko radarski podsistem z uporabo Dopplerjeve analize omogoča zaznavo premikajočih se objektov neodvisno od svetlobnih pogojev.

Radarski del vključuje Gaussovo glajenje, odštevanje referenčnega ozadja, pragovno segmentacijo ter gostotno gručenje z algoritmom DBSCAN, kar omogoča stabilno in robustno oceno razdalje in hitrosti tarče. Eksperimentalni rezultati potrjujejo, da kombinacija obeh pristopov poveča zanesljivost zaznave v primerjavi s posameznim sistemom, saj radar kompenzira omejitve vizualne detekcije pri slabših svetlobnih pogojih, medtem ko kamera omogoča klasifikacijo tipov objekta.

Možne nadaljnje izboljšave vključujejo razširitev sistema za zaznavo več tarč hkrati ter klasifikacijo različnih tipov brezpilotnih letalnikov na podlagi mikro-Dopplerjevih značilk.

Viri in literatura

- Eiadkaew S., Boonpoonga A., Phaebua K., Athikulwongse K., & Shutimarrungson N. (2022). Design and Experiment of FMCW Radar System for Drone Detection, *37th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC)*.
- Ester, M., Kriegel, H., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *KDD'96: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*. (str. 226 – 231).
- Jian, M., Lu, Z., & Chen, V. C. (2018). Drone detection and tracking based on phase-interferometric Doppler radar, *2018 IEEE Radar Conference (RadarConf18)*
- Obaid, L., Hamad, K., Al-Ruzouq R. (2025). State-of-the-art review of unmanned aerial vehicles (UAVs) and artificial intelligence (AI) for traffic and safety analyses: Recent progress, applications, challenges, and opportunities. *Transportation Research Interdisciplinary Perspectives*.
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2015). You Only Look Once: Unified, Real-Time Object Detection. *ArXiv*, <https://arxiv.org/abs/1506.02640> .
- Zhou, L. et.al. (2019). YOLO-RD: A lightweight object detection network for range doppler radar images. *IOP Conf. Ser.: Mater. Sci. Eng.* 563 042027.

Slovenski podjetniški sklad (SPS)

predstavlja ključno finančno podporno institucijo za mikro, mala in srednje velika podjetja v Sloveniji. Kot izvajalska institucija Ministrstva za gospodarstvo, turizem in šport sistematično krepi podjetniški ekosistem z dostopnimi in raznolikimi razvojnimi spodbudami.

V letu 2026 bo za različne razvojne spodbude namenil približno **159 milijonov evrov**, s katerimi bo podprl **okoli 3.000 podjetniških projektov** in s tem pomembno prispeval k nadaljnjemu razvoju slovenskega gospodarstva.

SPS svojo vlogo v okolju izvaja skozi 5 operativnih vlog:

Operativna vloga 1

FINANČNE SPODBUDE za MSP-je, start-upe in scale-upte

SPS podpira podjetja v vseh fazah razvoja – od zagona do rasti ter širitve na tuje trge. S širokim naborom finančnih spodbud podjetjem omogoča hitrejši razvoj, večjo konkurenčnost in lažji dostop do kapitala.

Podjetjem zagotavlja dostop do ugodnih finančnih virov, kot so:

- nepovratna sredstva,
- povratne spodbude (garancije in krediti),
- lastniško financiranje.

SPS podjetja spremlja na njihovi razvojni poti ter jim pomaga graditi stabilno poslovanje in uspešno uresničevati ambiciozne cilje doma in v tujini.

Operativne vloge



Operativna vloga 2

SPECIFIČNE SPODBUDE za razvoj znanja in kompetenc

Poleg finančne podpore SPS podjetjem zagotavlja tudi celovite vsebinske programe, ki krepijo poslovne in razvojne kompetence podjetnikov. Na voljo so:

- vavčerji,
- mentorstva,
- usposabljanja in povezovanja,
- svetovanja ter mreženja.

S temi programi SPS podjetjem pomaga razvijati znanja, izboljšati konkurenčnost in pospešiti rast na domačem in tujih trgih.

Operativna vloga 3

RAZVOJNA PARTNERSTVA za dvig inovativne usmerjenosti in trajnosti

SPS razvojna partnerstva snuje kot strateško povezovanje podjetij, raziskovalnih institucij in mednarodnih organizacij za spodbujanje inovativnosti, trajnostne rasti in globalne konkurenčnosti MSP, start-upov in scale-upov.

Partnerstva oblikuje glede na potrebe podjetij – kot vertikalne pospeševalniške programe na ključnih področjih (zdravje, pametna proizvodnja, biotehnologija, vodikove tehnologije) ali kot čezmejne pospeševalnike za vstop na tuje trge (ZDA, UAE).

S povezovanjem na nacionalni ravni ter nadgradnjo v mednarodne mreže podjetjem omogoča dostop do znanja, kompetenc, novih trgov in razvojnih priložnosti.

Operativna vloga 4

Promocija podjetništva in inovativnosti

SPS z različnimi programi, dogodki in pobudami promovira podjetništvo, inovativnost in ustvarjalnost – s ciljem, da spodbudimo razvoj uspešnih, trajnostno naravnanih podjetij, ki prispevajo k rasti slovenskega gospodarstva.

Operativna vloga 5

Učinkovita javna finančna institucija

Kot učinkovita javna institucija se SPS dosledno ravna po načelih enostavnosti, preglednosti in transparentnosti ter pri svojem delovanju upošteva ESG načela (okoljska, socialna in upravljavska). S strateškim usmerjanjem spodbud podpira trajnostni razvoj, digitalno preobrazbo, krožno gospodarstvo in družbeno odgovorno poslovanje. Ob tem aktivno soustvarja slovenski podjetniški ekosistem ter prispeva k uresničevanju evropskih ciljev na področju spodbujanja podjetništva, raziskav, razvoja in zaposlovanja.

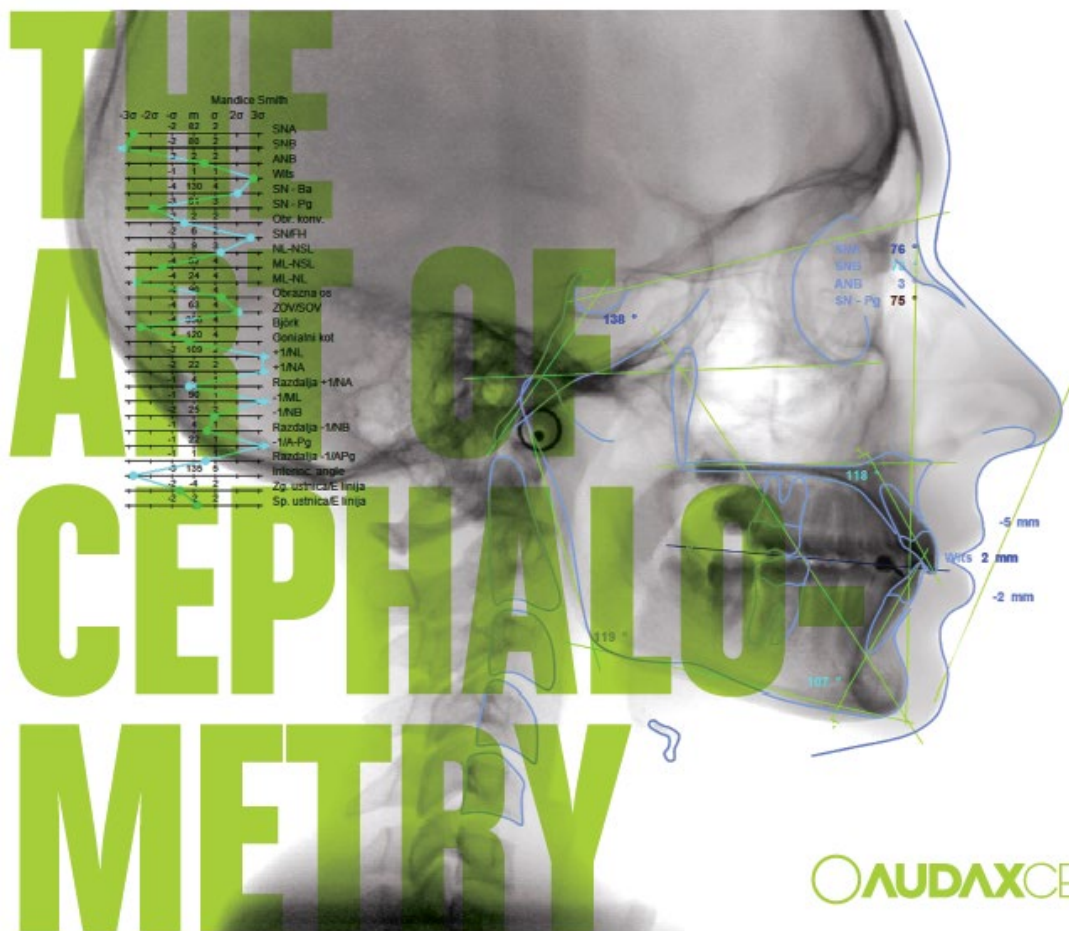
Slovenski podjetniški sklad – zanesljiv partner podjetjem na poti rasti in uspeha.

Kontaktne podatke

Slovenski podjetniški sklad
Ulica kneza Koclja 22, 2000 Maribor
Tel: 02/234 12 60
info@podjetniskisklad.si
www.podjetniskisklad.si

e-novice: priključite se v bazo in redno spremljajte informacije.





AUDAXCEPH®

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.

www.audaxceph.com

info@audaxceph.com

AUDAX d.o.o. - Termički park 18 - SI-1000 Ljubljana
Tel: +386 1 2500 40 50 • Fax: +386 1 423 47 53

ROSUS 2026 – RAČUNALNIŠKA OBDELAVA SLIK IN NJENA UPORABA V SLOVENIJI 2026: ZBORNİK 20. STROKOVNE KONFERENCE

BOŽIDAR POTOČNIK (UR.)

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

ROSUS 2026 – Računalniška obdelava slik in njena uporaba v Sloveniji 2026 je strokovna računalniška konferenca, ki jo od leta 2006 naprej vsako leto organizira Inštitut za računalništvo, Fakulteta 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, zabavišstva in sorodnih področij. Zbornik konference ROSUS 2026 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.4.2026](https://doi.org/10.18690/um.feri.4.2026)

ISBN
978-961-299-111-1

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

DOI
[https://doi.org/
10.18690/um.feri.4.2026](https://doi.org/10.18690/um.feri.4.2026)

ISBN
978-961-286-960-1

Keywords:

computer image processing,
machine vision,
biomedicine,
industrial applications,
knowledge transfer

ROSUS 2026 – COMPUTER IMAGE PROCESSING AND ITS APPLICATION IN SLOVENIA 2026: PROCEEDINGS OF THE 20TH 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

ROSUS 2026—Computer image processing and its application in Slovenia 2026 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 2026 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/rosus2026>
rosus.feri@um.si

