

Zbornik strokovne konference

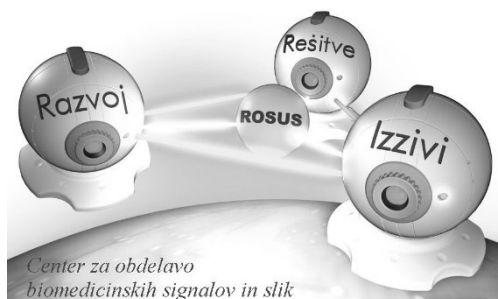
ROSUS 2023

Računalniška obdelava slik in
njena uporaba v Sloveniji 2023

Urednik:
Božidar Potočnik



Maribor, 23. marec 2023



ROSUS 2023 - Računalniška obdelava slik in njena uporaba v Sloveniji 2023

Zbornik 17. strokovne konference

Urednik
Božidar Potočnik

Marec 2023

Naslov <i>Title</i>	ROSUS 2023 - Računalniška obdelava slik in njena uporaba v Sloveniji 2023 <i>ROSUS 2023 - Computer image processing and its application in Slovenia 2023</i>
Podnaslov <i>Subtitle</i>	Zbornik 17. strokovne konference <i>Proceedings of the 17th Professional Conference</i>
Urednik <i>Editor</i>	Božidar Potočnik (Univerza v Mariboru, Fakulteta za elektrotehniko, računalništvo in informatiko)
Recenzija <i>Review</i>	Aleš Holobar (Univerza v Mariboru, Fakulteta za elektrotehniko, računalništvo in informatiko) Simon Dobrišek (Univerza v Ljubljani, Fakulteta za elektrotehniko)
Tehnični urednik <i>Technical editor</i>	Jan Perša (Univerza v Mariboru, Univerzitetna založba)
Oblikovanje ovitka <i>Cover designer</i>	Gašper Sedej (Univerza v Mariboru, Fakulteta za elektrotehniko, računalništvo in informatiko)
Grafične priloge <i>Graphic material</i>	Avtorji prispevkov, 2023
Konferenca <i>Conference</i>	Strokovna konferenca ROSUS 2023: Računalniška obdelava slik in njena uporaba v Sloveniji 2023
Datum in kraj konference <i>Conference date & location</i>	23. marec 2023, Maribor, Slovenija
Programski odbor <i>Program committee</i>	Aleš Holobar, Damjan Zazula, Boris Cigale, Marjan Mernik, Peter Peer, Tomaž Tomažič, Matjaž Colnarič, Slobodan Ribarič, Vojko Flis, Saša Divjak, Stanislav Kovačič, Zdravko Kačič, Aleš Leonardis, Simon Dobrišek, Vitomir Štruc, Franc Solina, Andrej Šostarič, 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, Filip Urh, Aleš Holobar, Božidar Potočnik
Založnik <i>Published by</i>	Univerza v Mariboru Univerzitetna založba Slomškov trg 15, 2000 Maribor, Slovenija https://press.um.si , zalozba@um.si
Izdajatelj <i>Issued by</i>	Univerza v Mariboru Fakulteta za elektrotehniko, računalništvo in informatiko Koroška cesta 46, 2000 Maribor, Slovenija https://feri.um.si , feri@um.si

Izdaja
Edition Prva izdaja

Vrsta publikacije
Publication type E-knjiga

Dostopno na
Available at <https://press.um.si/index.php/ump/catalog/book/769>

Izdano
Published Maribor, marec 2023



© Univerza v Mariboru, Univerzitetna založba Besedilo/ *Text*

University of Maribor, University Press

© Avtorji in Potočnik, 2023

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 (17 ; 2023 ; Maribor)

ROSUS 2023 [Elektronski vir] : računalniška obdelava slik in njena uporaba v Sloveniji 2023 : zbornik 17. strokovne konference : [23. marec 2023, Maribor, Slovenija] / urednik Božidar Potočnik. - 1. izd. - E-knjiga. - Maribor : Univerza v Mariboru, Univerzitetna založba, 2023

Način dostopa (URL): <https://press.um.si/index.php/ump/catalog/book/769>

ISBN 978-961-286-720-1 (PDF)

doi: 10.18690/um.feri.4.2023

COBISS.SI-ID 145392131

ISBN 978-961-286-720-1 (pdf)
978-961-286-721-8 (mehka vezava)

DOI <https://doi.org/10.18690/um.feri.4.2023>

Cena
Price Brezplačni izvod

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

Citiranje Potočnik, B. (ur.). (2023). *ROSUS 2023 - Računalniška obdelava*
Attribution *slik in njena uporaba v Sloveniji 2023: zbornik 17. strokovne konference*
Maribor: Univerzitetna založba. doi: 10.18690/um.feri.4.2023

POKROVITELJI

Audax

**SLOVENSKI
PODJETNIŠKI
SKLAD**



IEEE
Slovenska sekcija



SDRV
Slovensko društvo za razpoznavanje vzorcev
Slovenian Pattern Recognition Society

ROSUS 2023
<http://rosus.feri.um.si>

Kazalo

Predgovor	
<i>Foreword</i>	1
Božidar Potočnik	
VABLJENA PREDAVANJA	3
Strojni vid, ključna omogočitevna tehnologija v inteligentnih senzorjih podjetja SICK	
<i>Machine Vision, Key Enabling Technology in Sick Intelligent Sensors</i>	5
Jure Skvarč	
Kontrola pristopa s čitalniki prstnih odtisov v zasebnih objektih	
<i>Access Control with Fingerprint Readers in Private Objects</i>	15
Tadej Weber	
Identity Generation with Deep Generative Models	
<i>Generiranje identitet z globokimi generativnimi modeli</i>	25
Nunzio Alexandro Letizia	
Agricultural Field Delineation Using Satellite Imagery	
<i>Določevanje meja kmetijskih polj in uporabo satelitskih posnetkov</i>	31
Matej Batič, Jan Geršak, Matic Lubej, Žiga Lukšič, Nika Oman Kadunc, Devis Peressutti, Nejc Vesel, Sara Verbič	
DOPOLDANSKA IN POPOLDANSKA SEKCIJA	
Industrijske aplikacije, medicinske in biomedicinske aplikacije, drugo, študentske aplikacije	45
Modeliranje dinamike pečenja za določanje stanja pečenja z mrežami LSTM	
<i>Modeling Baking Dynamics for Determinining Baking State Using LSTM Networks</i>	47
Nina Sedej, Gregor Koporec, Janez Perš	
Zaznavanje oslonilnih točk aortne zaklopke v CT slikah s postopkom globokega učenja	
<i>Deep Learning-based Aortic Valve Landmark Detection in CT Images</i>	61
Luka Škrlj, Matija Jelenc, Franjo Pernuš, Tomaž Vrtovec	
Analiza geometrije lijakastih pasti v pesku	
<i>Analysis of Geometry of Pit-fall Traps in Sand</i>	69
Tina Klenovšek, Dušan Devetak, Jan Podlesnik	

Adapting VET Education to Labor Market Needs with Focus on Artificial Intelligence and Computer Vision	
<i>Prilagajanje poklicnega izobraževanja in usposabljanja potrebam trga dela z osredotočanjem na umetno inteligenco in računalniški vid</i>	77
Žiga Emeršič, Gregor Hrastnik, Nataša Meh Peer, Peter Peer	
Generiranje sintetičnih slik iz 3D modelov predmetov za učenje segmentacijskih modelov v industrijskem okolju	
<i>3D-model-based Rendering of Synthetic Images For Training Segmentation Models in an Industrial Environment</i>	87
Matic Fučka, Marko Rus, Jakob Božič, Danijel Skočaj	
Napovedovanje obrazov otrok z GAN metodo	
<i>Predicting Children Faces with GAN Method</i>	105
Nejc Šuklje, Luka Šajn	
Detekcija puščic pri klasičnem pikadu	
<i>Arrow Detection for Classic Darts</i>	115
Matic Zgonc, Borut Batagelj	
Unexploded Ordnance Detection on UAV Thermal Images by using YOLOv7	
<i>Detekcija neeksplodiranih ubojnih sredstev na termalnih slikah UAV s pomočjo YOLOv7</i>	125
Milan Bajjić, Božidar Potočnik	

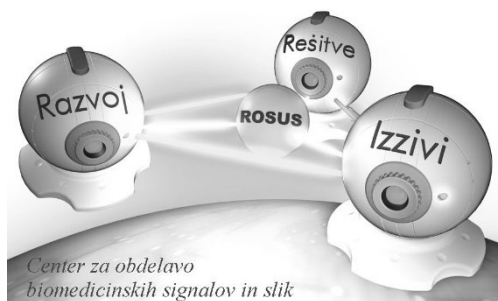
Spoštovani!

Po šestnajstih konferencah ROSUS 2006–2022 želimo s konferenco ROSUS 2023 nadaljevati s promoviranjem pomembnosti ekonomske koristi računalniške obdelave slik na področjih industrije, biomedicine in drugih poslovnih procesov. Vezi, ki smo jih na prejšnjih konferencah stkali med raziskovalci, razvijalci, ponudniki rešitev ter uporabniki računalniške obdelave slik v slovenskem prostoru, želimo še dodatno okrepiti, ob tem pa nuditi tudi 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 aplikacije računalniške obdelave slik nudijo nove priložnosti za uveljavitev na trgu visokih tehnologij. Seveda ne smemo pozabiti 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 še s pristrčno zahvalo Javnemu skladu Republike Slovenije za podjetništvo, ki v okviru konference ROSUS 2023 predstavlja zanimive finančne instrumente za spodbujanje prenosa tehnoloških rešitev v podjetniško sfero. Iskreno se zahvaljujemo tudi podjetju Audax za sprejem pokroviteljstva konference. Izpostaviti želimo še medijske pokrovitelje revijo IRT3000, revijo Svet elektronike, revijo Avtomatika + elektronika in revijo Ventil, ki so intenzivno promovirali konferenco ROSUS 2023 ter pomen strojnega vida v slovenskem prostoru.

Božidar Potočnik
predsednik konference
ROSUS 2023



VABLJENA PREDAVANJA

STROJNI VID, KLJUČNA OMOGOČITVENA TEHNOLOGIJA V INTELIGENTNIH SENZORJIH PODJETJA SICK

JURE SKVARČ

SICK d.o.o., Ljubljana, Slovenija
jure.skvarc@sick.si

Sinopsis SICK je eden vodilnih svetovnih ponudnikov rešitev za senzorske aplikacije v industrijskem sektorju. Podjetje s sedežem v Waldkirch im Breisgau je leta 1946 ustanovil Dr.-Ing. e. h. Erwin Sick. SICK ima več kot 11.000 zaposlenih po vsem svetu. Razvojna skupina, Vision Lab Slovenija je od avgusta 2022 del globalne razvojne skupine, ki deluje na področju avtonomne zaznave. Industrijska avtomatizacija se spreminja izjemno hitro. Na čelu razvoja najsodobnejše senzorske tehnologije SICK danes ponuja rešitve, ki so pripravljene na izzive prihodnosti – z inteligentnimi senzorji, ki zbirajo podatke, jih vrednotijo v realnem času, se prilagajajo svojemu okolju in komunicirajo v omrežje. V prispevku so podrobneje predstavljene rešitve, ki temeljijo na tehnologijah strojnega vida. Poudarjeni so pristopi, ki prek programske opreme podpirajo koncepte modularnosti in fleksibilnosti in posledično uporabnikom omogočajo visoko stopnjo prilagodljivosti rešitev specifičnim potrebam.

Ključne besede:
SICK,
inteligentni senzor,
strojni vid,
inovativnost,
programska
oprema

MACHINE VISION, KEY ENABLING TECHNOLOGY IN SICK INTELLIGENT SENSORS

JURE SKVARČ

SICK d.o.o., Ljubljana, Slovenia
jure.skvarc@sick.si

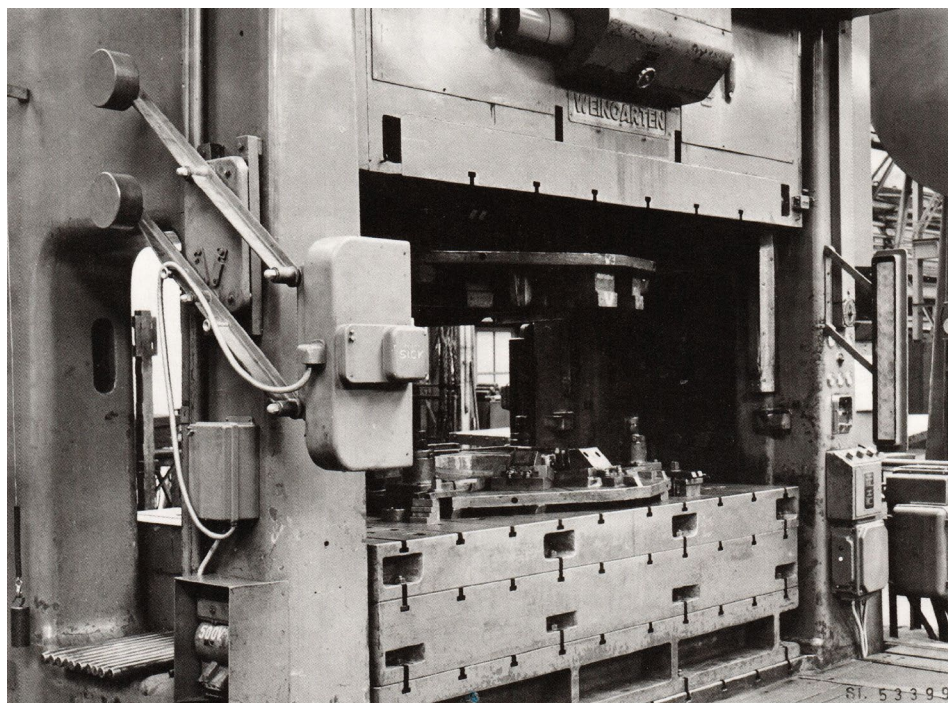
Abstract SICK is one of the world's leading suppliers of solutions for sensor applications in the industrial sector. Headquartered in Waldkirch im Breisgau, Germany, the company was founded in 1946 by Dr.-Ing. e. h. Erwin Sick. SICK has more than 11,000 employees worldwide. As of August 2022, the development group, Vision Lab Slovenia, is part of a global development team working in the field of autonomous perception. Industrial automation is changing at an extremely fast pace. As an innovation leader at the forefront of the development of state-of-the-art sensor technology, SICK today offers solutions that are ready for the challenges of the future - with intelligent sensors that collect data, evaluate it in real time, adapt to their environment and communicate in the network. In this paper, solutions based on image processing technologies are presented in more detail. Approaches are highlighted that support modularity and flexibility concepts through software, giving users a high degree of flexibility in solutions for specific requirements.

Keywords:

SICK,
intelligent sensor,
machine vision,
innovation,
software

1 Uvod

SICK je globalno priznan ponudnik industrijskih senzorskih rešitev in je bil ustanovljen v Nemčiji leta 1946. Podjetje zaposluje 11.000 ljudi v več kot 50 državah in se močno osredotoča na raziskave in razvoj. Leta 1952 je SICK na trg ponudil prvo svetlobno zaveso za preprečevanje nesreč in z inovacijo kmalu dosegel komercialni preboj. S časom se je portfelj izdelkov podjetja razširil tako, da vključuje različne senzorje za varnost, avtomatizacijo in kontrolo procesov.



Slika 1: Weingarten preša z vgrajeno SICK varnostno zaveso

Vir: (Fink 2004)

Zavezanost podjetja k raziskavam in razvoju ter osredotočenost na inovacije so podjetju omogočili ohranitev položaja vodilnega proizvajalca. Podjetje je usmerjeno k uporabnikom in tesno sodeluje s svojimi kupci, da bi razumelo uporabniške zahteve in zagotovilo prilagojene rešitve. Leta 2022 je družba SICK ustanovila Vision Lab Slovenija, skupino, ki se ukvarja z razvojem rešitev na področju

avtonomne zaznave. Pri tem uporabljajo najnovejše tehnologije na področju strojnega vida in strojnega učenja.

1.1 Tehnologije strojnega vida in njihova uporaba v industrijski avtomatizaciji

V prispevku izpostavljamo tehnologije strojnega vida in njihovo uporabo za industrijsko avtomatizacijo. Strojni vid je inženirsko področje, ki se ukvarja z uporabo kamer in računalniških algoritmov, z namenom, da strojem zagotovi sposobnost zajema in razumevanje vizualnih podatkov. Ta tehnologija postaja vse pomembnejša pri industrijski avtomatizaciji, saj strojem omogoča obdelavo in analizo slik v realnem času ter sprejemanje odločitev na podlagi informacij, ki jih sistemi ustvarjajo.

V članku bomo raziskali različne uporabe tehnologij strojnega vida v industrijski avtomatizaciji, vključno z nadzorom kakovosti, logistiki, spremljanju procesov in robotiko. Preučili bomo, kako se te tehnologije uporabljajo za izboljšanje učinkovitosti v proizvodnih procesih ter za skrajšanje zastojev in nižanje izmeta.

Predstavili bomo primere uspešnih implementacij tehnologij strojnega vida na področju industrijske avtomatizacije in nakazali trende za prihodnji razvoj na tem področju.

2 Spreminjanje pokrajine industrijske avtomatizacije

Industrijska avtomatizacija se hitro spreminja zaradi napredka v tehnologiji, spreminjanja zahtev uporabnikov in potrebe po učinkovitosti in konkurenčnosti. Tradicionalni pristop neodvisnih strojev v proizvodnji se umika konceptu digitalizirane proizvodnje, medsebojno povezanim napravam, podatkovni pristopom z industrijo 4.0 na čelu. Takšen pristop omogoča višjo stopnjo avtomatizacije, komunikacijo med stroji in ustvarjanje pametnih tovarn. Povpraševanje po prilagojenih izdelkih je vodilo v razvoj modularnih in razširljivih avtomatiziranih sistemov, medtem ko masovni podatki in napredna orodja analitike omogočajo sprejemanje odločitev v dejanskem času ter preventivno vzdrževanje. Stalen razvoj na področju robotike in drugih naprednih tehnologij vodi do prožnejših in učinkovitejših proizvodnih sistemov. Ta razvoj dogodkov ustvarja

fleksibilne, modularne, podatkovno vodene učinkovite proizvodne sisteme, ki se lahko hitro odzovejo na tržne razmere in potrebe strank.

3 Tehnologije strojnega vida v podjetju SICK

Tehnologije strojnega vida, vgrajene v produkte podjetja SICK omogočajo avtomatizacijo zajema podatkov in analizo vizualnih informacij v industrijskih in drugih aplikacijah.



Slika 2: Primeri izdelkov podjetja SICK, ki temeljijo na tehnologijah strojnega vida

Vir: lasten.

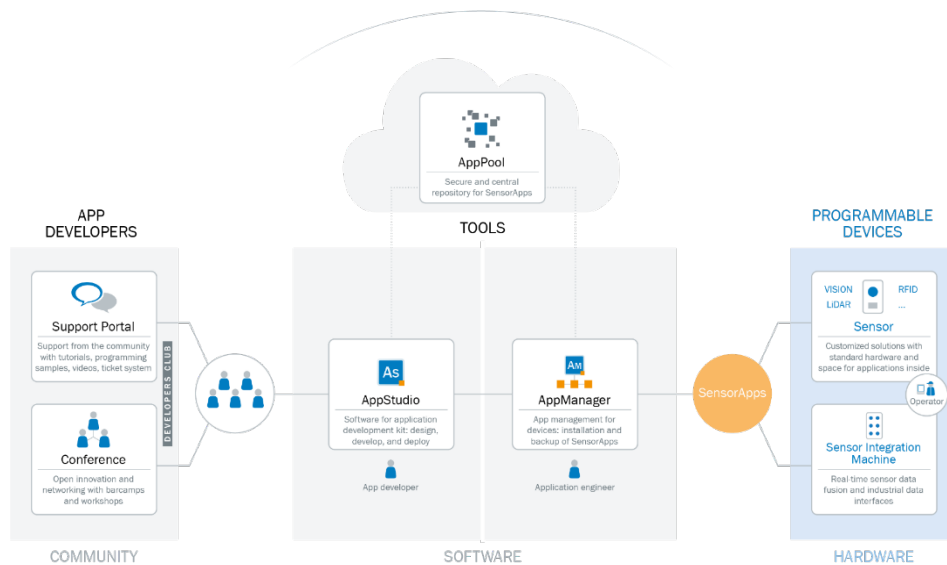
Izdelki se uporabljajo v številnih panogah, vključno v proizvodnji, logistiki in robotiki. Tipični izdelki podjetja SICK s tehnologijami strojnega vida so:

Inteligentni senzorji: SICK ponuja vrsto inteligentnih senzorjev (1D, 2D, 3D), ki združujejo zmogljivosti zajema vizualnih podatkov z zmogljivostmi obdelave podatkov in komunikacije. Ti senzorji so zasnovani tako, da zajemajo in analizirajo vizualne informacije znotraj naprave, v dejanskem času, kar omogoča izvedbo številnih rešitev. Naprave so zgrajene okoli širokega nabora senzorjev, od črno-belih CMOS senzorjev do več-modalnih 3D senzorjev, ki npr. hkrati ponujajo informacijo o oddaljenosti, barvi in raztrosu.

Kamere: SICK ponuja vrsto visokokakovostnih kamer (2D, 3D), ki so namenjene uporabi v zahtevnih industrijskih aplikacijah. Ena od posebnosti so kompaktne 3D kamere, ki delujejo na principu časa preleta (angl.: time of flight).

Laserski senzorji za zaznavanje razdalje: Senzorji za lasersko merjenje razdalje so namenjeni merjenju razdalj in zaznavi oddaljenosti v širokem razponu aplikacij. Ti senzorji so zelo točni, robustni in lahko delujejo v različnih zahtevnih okoljih, vključno v industrijskem okolju in na prostem.

Programska oprema: Podjetje SICK ponuja vrsto programskih rešitev, ki so zasnovane tako, da podpirajo in nadgrajujejo strojne komponente. Te programske rešitve so zelo prilagodljive in razširljive in so zasnovane tako, da izpolnjujejo specifične potrebe vsakega kupca. Vključujejo zaključene rešitve, kot so varnostni senzori ali zaznavanje predmetov za robotske aplikacije. Popolnoma nastavljive aplikacije, kot je npr. NOVA, omogočajo izvedbo zmerno zahtevnih rešitev kontrole kakovosti brez programiranja. Za napredne uporabnike obstajajo orodja, kot sta AppDesigner in AppManager, ki razvijalcem omogočata ustvarjanje lastnih aplikacij strojnega vida, vključno z lastnim uporabniškim vmesnikom. Nabor orodij (dTag, dStudio, dLearner) tudi neizkušenim uporabnikom omogoča razvoj lastnih modelov globokega učenja, ki jih je mogoče nadalje uporabljati v aplikacijah. SICK povezuje močno skupnost uporabnikov in razvijalcev, ki nenehno izzivajo obstoječe rešitve in spodbujajo nadaljnji razvoj.



Slika 3: Tri ključne komponente ekosistema AppSpace podjetja SICK

Vir: lasten.

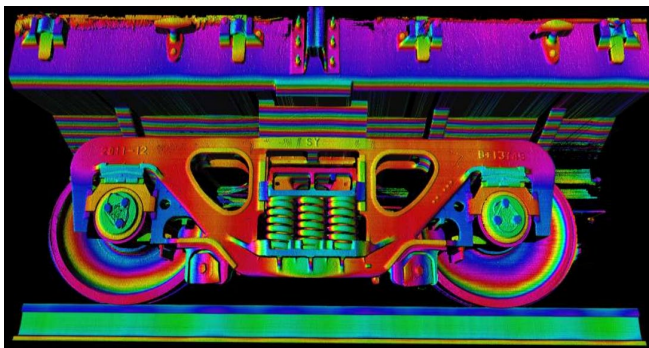
Podatke, ki jih ustvarijo senzorske aplikacije, ki se izvajajo v napravah SICK, lahko še dodatno uporabljajo aplikacije v oblaku, kot so SensorStudio, Field Analytics, SARA in Asset Hub. Te aplikacije v oblaku omogočajo napredno analizo, vizualizacijo in upravljanje podatkov, ki jih ustvarjajo aplikacije senzorjev, kar uporabnikom omogoča sprejemanje odločitev in optimizacijo svojih industrijskih procesov. S tem bogatijo zmogljivosti tehnologij strojnega vida SICK in

uporabnikom zagotavljajo kakovostno in celovito rešitev za industrijsko avtomatizacijo.

4 Primeri uporabe

Tehnologije strojnega vida SICK se uporabljajo v številnih aplikacijah v realnem svetu, vključno z nadzorom kakovosti, navigacijo, zagotavljanje varnosti in logistiko.

Nadzor kakovosti se nanaša na uporabo teh tehnologij za pregled izdelkov in preverjanje kakovosti med proizvodnim postopkom. Tehnologije strojnega vida SICK uporabljajo napredne algoritme za odkrivanje napak in odstopanj od zadanih standardov, kar zagotavlja, da le visokokakovostni izdelki dosežejo trg. Z rešitvami za hitro 3D pregledovanje in napravami, ki omogočajo strojno učenje na napravi (inteligentni senzor) je SICK pred vsemi svetovnimi konkurenti.



Slika 4: Hiter zajem 3D podatkov vlakovne kompozicije med gibanjem

Vir: lasten.

Navigacija in zagotavljanje varnosti se nanašata na uporabo tehnologij strojnega vida za varno in učinkovito gibanje vozil in ljudi v različnih razmerah. Tehnologija SICK LIDAR v dejanskem času ustvarja informacije o okolju, kar avtonomnim vozilom in robotom omogoča varno in natančno gibanje. LIDAR sistemi zagotavljajo sočasne podatke v več prerezih in lahko robustno delujejo v zahtevnih okoljih. Številne naprave so certificirane tudi kot varnostne naprave.

Na področju *logistike* se tehnologije uporabljajo v skladiščih, letališčih in poštah za racionalizacijo poslovanja in povečanje učinkovitosti. Proizvodi podjetja SICK se uporabljajo za sledenje in lokalizacijo paketov in zabojnikov ter avtomatizacijo obdelave in prevoza blaga. Tipičen primer so sodobni logistični centri. Čitalci 1D in 2D kod visoke ločljivosti lahko optično preberejo obsežno vidno polje in v nekaj milisekundah prilagodijo optiko glede na oddaljenost predmetov.

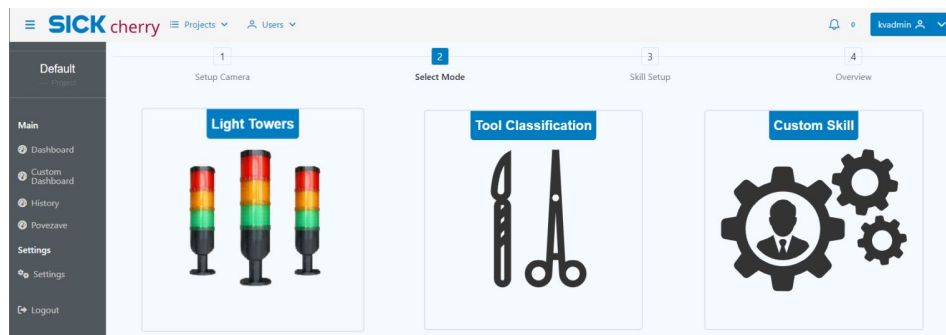


Slika 5: Branje 1D in 2D kod na vseh straneh paketov

Vir: lasten.

Vision Lab Slovenija razvija VIVIAN, kognitivnega asistenta, ki je zasnovan tako, da uporabnikom pomaga pridobiti dragocen vpogled v proizvodni proces. Ponuja več sposobnosti in se lahko samodejno prilagodi poljubnemu proizvodnem procesu z opazovanjem, zbiranjem in analizo vizualnih informacij. Za razliko od tradicionalnih metod (npr. branje podatkov iz PLC krmilnikov), je VIVIAN neinvazivna rešitev. Posnema človeka, vendar brez tveganja za človeške napake. Z VIVIAN lahko bolje razumemo proizvodni proces in sprejemamo odločitve o izboljšanju učinkovitosti in produktivnosti. Poleg tega lahko VIVIAN uporabnike opozori na morebitna odstopanja v delovanju, kar jim omogoča, da jih odpravijo še preden bi lahko nastale motnje v proizvodnji.

VIVIAN ustvarja dodano vrednost tako v starejših proizvodnih okoljih kot v najsodobnejših avtomatiziranih procesih z naprednim spremljanjem in analizo delovanja, vizualizacijo podatkov in obveščanjem. Rezultat so učinkovitejši procesi in višja produktivnost



Slika 6: Pogled na uporabniški vmesnik rešitve VIVIAN

Vir: lasten.

5 Zaključek

SICK je vodilni ponudnik tehnologij strojnega vida in senzorskih rešitev z dolgo zgodovino inovacij in močnim poudarkom na raziskavah in razvoju. Inteligentni senzori in programske rešitve podjetja omogočajo pridobivanje, analizo in komunikacijo v realnem času, kar vodi k višji produktivnosti v različnih panogah. Na hiter razvoj industrijske avtomatizacije se SICK odziva z rešitvami strojnega vida, ki ponujajo modularnost in fleksibilnost in omogočajo prilagoditev specifičnim potrebam uporabnikov. Osredotočenost podjetja na inovacije, v kombinaciji s prilagodljivim in razširljivim obsegom izdelkov, podjetju omogoča, da izpolnjuje razvijajoče potrebe industrije in ostane pred konkurenco. Z ustanovitvijo skupine Vision Lab Slovenija, SICK še dodatno vpliva v področje avtonomne zaznave in si s tem zagotavlja konkurenčno prednost na področju senzorski tehnologiji za tovarne prihodnosti.

Literatura

Fink, E. (2004). Erwin Sick Inventor and Entrepreneur, Burger Druck e.K., Waldkirch-Kollnau

KONTROLA PRISTOPA S ČITALNIKI PRSTNIH ODTISOV V ZASEBNIH OBJEKTIH

TADEJ WEBER

NAVKOM d.o.o., Brezovica, Slovenija
tadej@navkom.si

Sinopsis Zasnova čitalnikov prstnih odtisov, ki se vgrajujejo in uporabljajo v zasebnih objektih (hiše in stanovanja) ima svojo specifiko predvsem v tem, da je potrebno upoštevati različne navade uporabnikov, starostne skupine uporabnikov, izpostavljenost vremenskim vplivom ter zahteve proizvajalcev vhodnih vrat glede načina vgradnje. S pravilnim informiranjem vseh deležnikov v procesu od proizvodnje vrat do končnega uporabnika, ter s pravimi nastavitvami algoritma, ki obdeluje zajete biometrične podatke, lahko zagotovimo tudi več kot 95 % uspešnih transakcij. Daleč največji izziv nam je predstavljalo informiranje in izobrazba uporabnikov, še posebej preden se je tovrstna tehnologija začela bolj množično pojavljati v mobilni telefoniji. Tako kot se je to zgodilo v mobilni telefoniji, postaja dandanes odklepanje vhodnih vrat s pomočjo biometrične prepoznavne standard.

Ključne besede:

čitalnik prstnih odtisov, varnost, biometrija, biometrični podatki, kontrola pristopa

ACCESS CONTROL WITH FINGERPRINT READERS IN PRIVATE OBJECTS

TADEJ WEBER

NAVKOM d.o.o., Brezovica, Slovenija
tadej@navkom.si

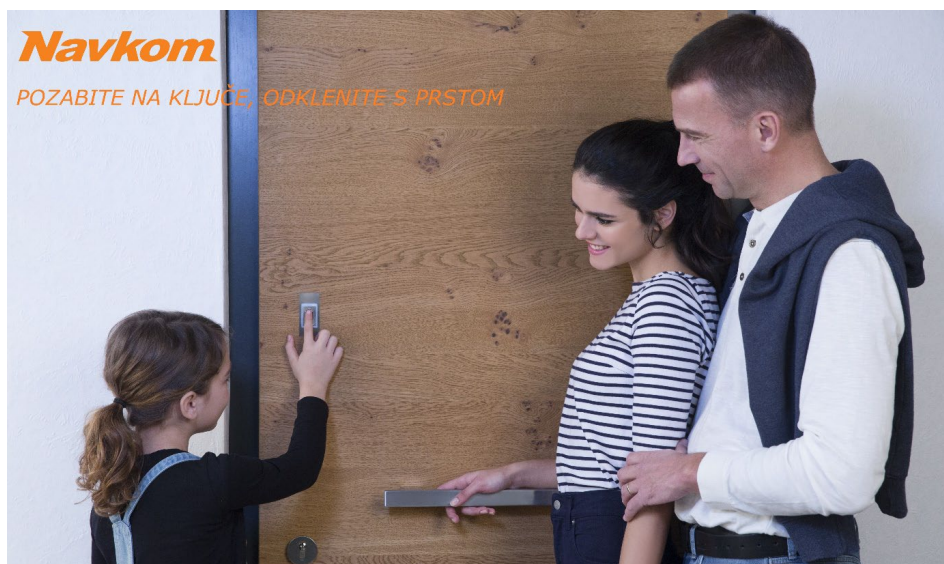
Abstract The design of fingerprint readers that are installed and used in private objects (houses and apartments) has its specificity, mainly in that it is necessary to take into account the different habits of users, age groups of users, exposure to weather influences and the requirements of entrance door manufacturers regarding the method of installation. By properly informing all participants in the process from door production to the end user, and with the correct settings of the algorithm that processes the captured biometric data, we can ensure more than 95% of successful transactions. By far the biggest challenge for us was informing and educating users, especially before this type of technology began to appear more widely in mobile telephony. Just as it happened in mobile telephony, unlocking the entrance doors with the help of biometric recognition is becoming the standard these days.

Keywords:
fingerprint reader,
security,
biometry,
biometric data,
access control

1 Uvod

Odklepanje vhodnih vrat s pomočjo biometrične identifikacije je poleg uporabe na mobilnih napravah druga najbolj razširjena korist te tehnologije med zasebnimi uporabniki. Naše podjetje se z razvojem in proizvodnjo tovrstnih naprav ukvarja od leta 2005 in v tem članku opisujem naše dosedanje izkušnje ter mnenje, kaj se bo dogajalo v bližnji prihodnosti.

Prav uporaba biometrične identifikacije na mobilnih napravah je v zadnjih 10 letih spodbudila masovno uporabo biometrije tudi pri odklepanju vhodnih vrat v zasebnih objektih, saj so uporabniki nekoliko omilili svoje predsodke glede uporabe biometrične kontrole pristopa in njene zanesljivosti. Ravno predsodek glede zanesljivosti tehnologije je naš glavni izziv.



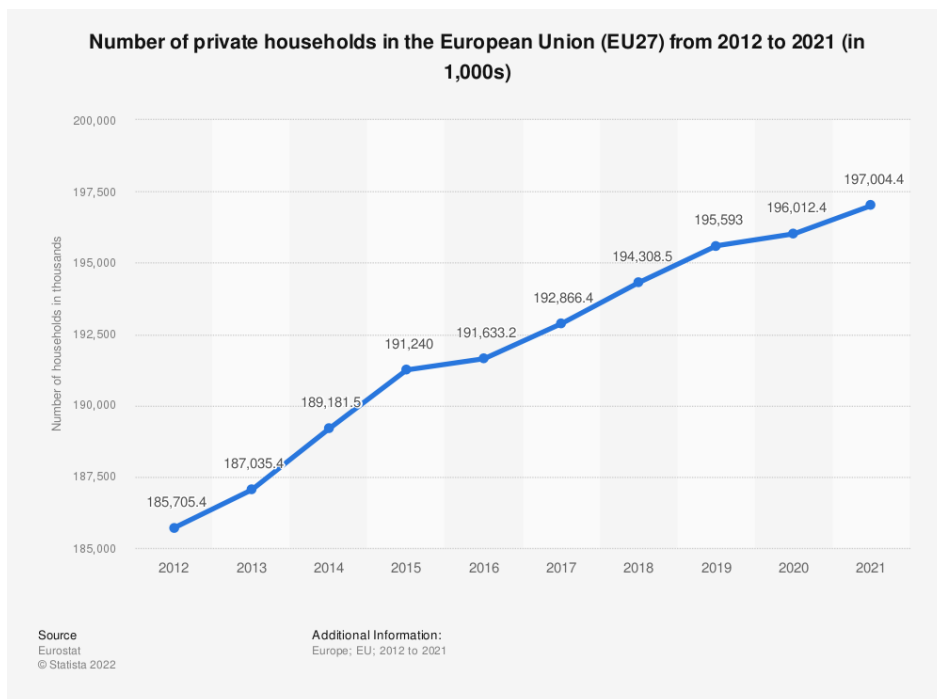
Slika 1: Pristop s čitalnikom prstnega odtisa

Vir: Arhiv NAVKOM d.o.o.

2 Metodologija

Če so leta 2010 le redki proizvajalci vhodnih vrat kot dodatno opremo ponujali biometrično kontrolo pristopa in se je s to tehnologijo opremilo manj kot 1 % vhodnih vrat, je danes situacija precej drugačna. Nekateri proizvajalci vhodnih vrat

opremijo že več kot 20 % vrat s čitalnikom prstnih odtisov in razmišljajo celo, da bi to postala standardna oprema vrat. Ob upoštevanju dejstva, da se samo v Evropi letno zgradi več kot milijon (povprečno 1,3 M) novih stanovanjskih hiš in naraščajočem deležu vrat, ki so opremljena s kontrolo pristopa, ocenjujemo, da bo tovrstna tehnologija šele v naslednjih desetih letih doživela svoj razcvet. Zasebni uporabniki zahtevajo vedno več udobja in odklepanje vrat brez ključa je ena od navad, ki se jo zelo hitro navadimo in potem brez tega težko shajamo. Zahteva se tudi več kontrole nad vhodom v hišo, tudi na daljavo, ter povezavo vseh sistemov hišne avtomatizacije, kar omogoča ne samo varnost in nadzor nad vstopanjem, temveč tudi prihranke energije. S povezavo na internet se naprave lahko posodablajo, opravlja se diagnostika in po potrebi servis na daljavo. V zadnjem času je vse večji poudarek na krožnem gospodarstvu, naprave se dizajnirajo tako, da je možno čim več sestavnih delov ponovno uporabiti, naprave se lažje servisirajo in posodablajo.



Slika 2: Število zasebnih objektov v EU med letoma 2012 in 2021

Vir: Eurostat, 2022

Zadnjih dvajset let, kolikor je minilo od prvih čitalnikov prstnih odtisov za odklepanje vhodnih vrat, ki so pri nas prišli na trg kot naprave za dodatno vgradnjo na obstoječa vrata, se je v tem segmentu veliko spremenilo. To je bilo obdobje uvajanja biometrične kontrole pristopa za zasebna vhodna vrata. V tem času so se sistemi prilagodili tako, da se čitalnik prstnih odtisov lahko vgradi neposredno v vratno krilo, v ročaj vhodnih vrat ali celo v kljuko in v cilindrični vložek/ključavnico na vhodnih vratih. To pomeni predvsem, da smo morali čitalnike čim bolj zmanjšati. Največji del elektronike je tako postal sam senzor, ki zajema podobo prstnega odtisa. Zaradi cenovnih omejitev se uporablja manj zmogljive procesorje, s katerimi obdelujemo zajete podatke, shranjujemo pa tako imenovane template, in sicer kot »minutiae« in relacije med njimi, ki zavzamejo malo spominskega prostora, zato je elektronsko vezje lahko manjše od senzorja.



Slika 3: »Minutiae«, ki označujejo specifične točke prstnega odtisa

Vir: Arhiv NAVKOM d.o.o.

Poleg čedalje bolj zmogljivih algoritmov za obdelavo zajetih biometričnih podatkov, pa se je v teh letih nekaj razvoja zgodilo tudi na samih senzorjih za zajem prstnega odtisa. Na voljo je veliko različnih senzorjev, ki temeljijo na različnih tehnikah (optični, kapacitivni, termični ...), vendar je pri napravah za kontrolo pristopa v zasebnih objektih potrebno najti pravo ravnovesje med ceno in zanesljivostjo naprave. Zato so se pri nas v začetku uporabljali večinoma termični senzorji podjetja *Atmel*®, preko katerih je potrebno prst potegniti (imenujemo jih senzorji na poteg), dobili smo zaporedje termičnih slik (temperatura izbokline in vdolbine vijuge

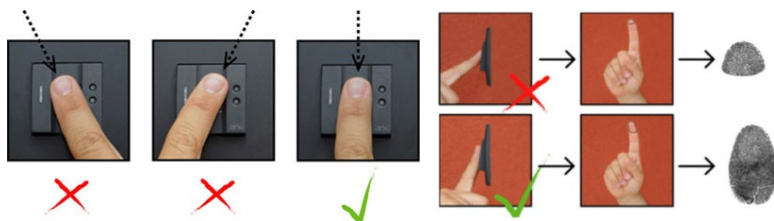
prstnega odtisa imata različni vrednosti), ki smo jih nato sestavili v prstni odtis. Leta 2009 ti senzorji niso bili več dobavljivi, zato se je večina proizvajalcev odločila za uporabo kapacitivnih senzorjev na poteg, ki so tudi bili cenovno bolj ugodni.



Slika 4: Čitalniki prstnih odtisov s senzorjem na dotik

Vir: Arhiv NAVKOM d.o.o.

Mi smo se takrat odločili za prehod na kapacitivne senzorje, na katere se prst prisloni (imenujemo jih senzorji na dotik). Ti senzorji so večji in tudi dražji, omogočajo pa bistveno bolj enostavno uporabo, kar je po naših pričakovanjih naletelo na pozitiven odziv uporabnikov. Znano je bilo, da ima velik del uporabnikov težavo s potegovanjem prsta po senzorju (ker je bilo potrebno potegniti po senzorju s pravo hitrostjo in optimalnim pritiskom), predvsem če je uporabnikov relief prstnega odtisa manj intenziven ali poškodovan.



Slika 5: Pravilna uporaba senzorja na poteg

Vir: Arhiv NAVKOM d.o.o.

Senzorje na dotik so v zadnjih letih nekoliko izboljšali, tako da smo lahko začeli uporabljati tudi takšne z nekoliko manjšo površino in imeli še vedno dobre rezultate, na voljo pa so vse manjši senzorji na dotik, celo z aktivno površino 8 x 8 mm, kar

pa pri tehnologiji, kjer ne shranjujemo slike prstnega odtisa, vodi v težave pri prepoznavanju prstnih odtisov, ki so manj intenzivni, poškodovani, z ekstremno suho kožo. Pri senzorjih z manjšo aktivno površino moramo zato odvzeti več zaporednih odčitkov istega prstnega odtisa v različnih pozicijah. To pa pri uporabnikih ni naletelo na pozitiven odziv. V kolikor zmanjšamo število odčitkov, po navadi so to trije zaporedni odčitki, seveda tvegamo kasnejše težave pri prepoznavi prstnega odtisa, zato se senzorji z manjšo aktivno površino sedaj kombinirajo s številčno tipkovnico, da v primeru težave s prepoznavo prstnega odtisa uporabnik lahko vtipka kodo, s katero odklene vrata. Glavno vodilo je namreč, da se prepreči situacijo, ko uporabnik ne more v lastno hišo. Težave s prepoznavo so sicer redke, vendar na tržišču bolj opazne. Ljudje smo namreč bolj dovzetni za negativne informacije o izdelkih kot pozitivne. Danes izbira prave velikosti oziroma aktivne površine senzorja v kombinaciji z dobrim algoritmom daje optimalne rezultate pri širšem krogu uporabnikov. V zadnjem času so tudi ostali proizvajalci tovrstnih izdelkov prepoznali prednosti senzorja na dotik, tako da večina novih izdelkov na trgu vsebuje te senzorje. V večini uporabljamo senzorje švedskega proizvajalca *Fingerprints cards*©.



Slika 6: Primer slabše kvalitete prstnega odtisa

Vir: Arhiv NAVKOM d.o.o.

Kombinacija tipkovnice in čitalnika prstnih odtisov pa postaja vse bolj popularna, ker nam omogoča tudi druge namene uporabe in sicer v objektih, ki se jih uporablja v turistične oziroma najemniške namene. Večina evropskih držav zakonsko omejuje

uporabo biometričnih podatkov, predvsem ko gre za tretje osebe. Do sedaj je večina čitalnikov prstnih odtisov v zasebnih objektih vgrajenih kot samostojna naprava, ki ni povezana z internetno povezavo. V nekaterih primerih se je s posebnimi vmesniki dalo čitalnik povezati s sistemom hišne avtomatizacije. Danes uvajamo čitalnike prstnih odtisov in tipkovnice, ki se lahko povežejo na internet in upravljajo oddaljeno, kar je predvsem prednost za tipkovnice, ki so vgrajene v turistične objekte. Prednost teh naprav je tudi odklepanje vrat na daljavo in upravljanje baze prstnih odtisov na daljavo.



Slika 7: BioTouchpad, kombinacija tipkovnice in čitalnika prstnega odtisa

Vir: Arhiv NAVKOM d.o.o.

Kljub temu, da je prepoznavna obraza lahko bolj učinkovita in udobna, ter se že uporablja tudi za kontrolo pristopa v zasebnih objektih, ostajajo čitalniki prstnih odtisov bolj popularna izbira predvsem zaradi nizke cene, majhnih dimenzij in enostavne vgradnje. V letih 2020 in 2021, je zaradi epidemije že kazalo, da bodo prevladale tehnologije brezstične kontrole pristopa. Vendar se je nato trend spet obrnil in čitalniki prstnih odtisov zaenkrat ostajajo v ospredju. Kljub temu verjamemo, da bo v prihodnosti prevladala prepoznavna obraza oziroma kakšen drug način prepoznave unikatnih človeških značilnosti s pomočjo kamere. Sčasoma bodo tudi ti sistemi cenovno bolj dostopni in zanesljivi. Trenutni izziv obstoječih naprav

za prepoznavo obraza ali očesa je v tem, da se moramo napravi precej približati in je ta nameščena na višini, ki ustreza povprečni višini odraslega človeka, pri tem imajo težave otroci, invalidi in ljudje višje rasti. Takšna naprava mora nujno delovati na večji razdalji (vsaj 2 m). Zelo verjetno kamere ne bo nujno potrebno vgrajevati v sama vrata in se bo lahko uporabljalo video nadzor okolice objekta. Pri uporabi zunanje kamere seveda ne vemo, ali oseba, ki smo jo prepoznali, res želi odpreti vhodna vrata, ali gre samo mimo in ima druge namene. Pri tem bo zagotovo v pomoč umetna inteligenca, ki se tudi že vpeljuje v hišne sisteme.

Proizvajalci vhodnih vrat so večinoma zelo konzervativni in zelo previdni pri uvajanju novih tehnologij. Vhodna vrata morajo - za razliko od mnogih drugih izdelkov - imeti daljšo življenjsko dobo in za vse, kar je vgrajeno v vrata se tudi pričakuje, da bo delovalo dlje časa. Popravila v garancijski dobi, ki je pri vhodnih vratih tudi 5 let in več, so večinoma relativno draga, ker se mora tehnik odpeljati do stranke in tam odpraviti napako. Uporabniki namreč ne morejo sami posegati v sestavne dele vhodnih vrat. Tudi zato so proizvajalci vhodnih vrat zelo previdni pri izbiri dobavitelja kontrole pristopa, ki se vgradi v vhodna vrata. Veliko vlogo igra torej sama zanesljivost naprave, nato enostavnost vgradnje in nenazadnje enostavno vzdrževanje. V primeru okvare mora biti zamenjava naprave čim bolj enostavna.

Proizvajalci vhodnih vrat bodo lahko po eni strani razbremenjeni vgradnje sistema kontrole pristopa v vhodna vrata in bo v tem primeru potrebno vgraditi samo motorni sistem zaklepanja. Po drugi strani pa imajo z vgradnjo dodatnega sistema kontrole pristopa tudi dodatne zasluge. Menimo, da se bodo naprave za prepoznavo biometričnih lastnosti še naprej vgrajevale tudi v vhodna vrata in bodo vse bolj povezane tudi z dodatnimi senzorji za nadzor delovanja vhodnih vrat, s katerimi se bo zagotavljalo nemoteno delovanje in nadzor.

Literatura

Eurostat, <https://ec.europa.eu/eurostat> (zadnje dostopno: 20. 2. 2023)

IDENTITY GENERATION WITH DEEP GENERATIVE MODELS

NUNZIO ALEXANDRO LETIZIA

PiktID, Klagenfurt, Austria
nunzio.letizia@piktid.com

Abstract Deep generative models have drawn the attention of the AI community in the last decade. The scalability of neural architectures helps solving multiple relevant problems, e.g., text-to-image generation, otherwise not addressable. In the context of image data privacy, the increasing amount of produced, shared, and stored images imposes new measures to protect personal identity information. At the same time, such protection mechanisms need to preserve the image quality. In this talk, we discuss how PiktID is using some recent deep learning-based techniques for protecting human identities in pictures. We show several examples, and we present interesting use-cases.

Keywords:

deep learning,
anonymization,
generative models,
image privacy,
data protection

GENERIRANJE IDENTITET Z GLOBOKIMI GENERATIVNIMI MODELI

NUNZIO ALEXANDRO LETIZIA

PiktID, Celovec, Avstria
nunzio.letizia@piktid.com

Sinopsis Globoki generativni modeli so v zadnjem desetletju pritegnili pozornost skupnosti AI. Razširljivost nevronske arhitekture pomaga pri reševanju več pomembnih problemov, npr. ustvarjanje besedila v sliko, ki sicer ni naslovljivo. V kontekstu zasebnosti slikovnih podatkov vse večja količina proizvedenih, deljenih in shranjenih slik nalaga nove ukrepe za zaščito informacij o osebni identiteti. Hkrati morajo takšni zaščitni mehanizmi ohraniti kakovost slike. V tem govoru razpravljamo o tem, kako PiktID uporablja nekatere nedavne tehnike, ki temeljijo na globokem učenju, za zaščito človeških identitet na slikah. Prikazujemo več primerov in predstavljamo zanimive primere uporabe.

Ključne besede:

globoko učenje,
anonimizacija,
generativni modeli,
zasebnost slike,
zaščita podatkov

1 Introduction

Anonymization is the process of removing personal identifiable information from data, in our case, images. In particular, we will mostly refer to sensitive data as the biometric information in the form of facial and body features.

PiktID aims at substituting biometric information available in images with new synthetic one not belonging to any human. The objective is to ensure security and enhance private protection, thus inhibiting re-identification of the original subjects from anonymized data.

However, to reach the final full anonymization goal, several technological and technical challenges need to be firstly defined, addressed and lastly solved.

2 Methodology

Two distinct classes of anonymization solutions have been developed in literature. Classical image processing-based and deep-learning based. The former solutions aim at detecting the sensitive subject and cover the information typically with filters such as Gaussian or bounding boxes applied on specific regions (that can also be defined via segmentation).

However, these techniques have the major flaw of being immediately visible to the human eye, thus significantly decreasing the quality, and of not preserving the statistical information that in some applications may be relevant (e.g., extracting statistical information of a shop's visitors being GDPR compliant) (Yang 2022).

The latter approach, instead, has recently received attention and some initial works based on generative models have been already proposed but only applied to faces (Zhongzheng, 2018). In particular, it is worth mentioning DeepPrivacy (Håkon, 2019) and CFA-NET (Tianxiang, 2021) since they both use Generative Adversarial Networks (GANs) (Goodfellow, 2014) and the second one also uses StyleGAN (Karras, 2020).



Figure 1: Examples from the current PiktID face anonymization tool. Quality is preserved.

Source: own.

Scaling GANs to hair and body anonymization is partially an unsolved task and it seems that diffusion models (Jonathan, 2020) is the most suitable generative model in this scenario. In fact, diffusion models have shown a more versatile generation process compared to GANs, enabling the generation of more complex and diverse images, thanks to the guidance of a language model. However, they still lack precision in details (high-frequency components) and the research community is working towards refinement of models such as Stable Diffusion.

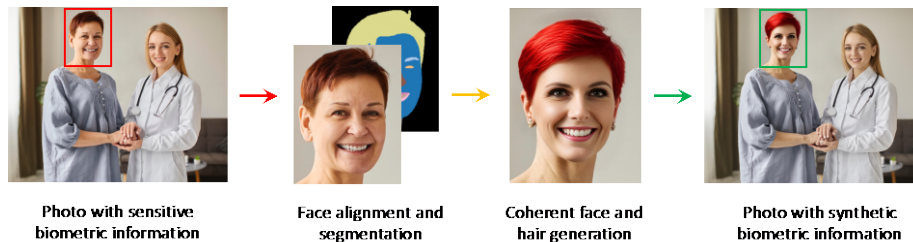


Figure 2: PiktID workflow for face and hair anonymization. The biometric information is extracted from the vulnerable subject and it is replaced with synthetic one.

Source: own.

When substituting faces, hair or the full body with synthetic biometric information, it often happens that a mismatch occurs and a background prediction block is needed. For instance, suppose to change the hairstyle in a way that the new hair is much shorter than the original one. The initially covered regions, that are now substituted, require a background prediction block that needs to understand the scene and coherently fill in the region based on the surroundings. Due to the variety of possible scenes, diffusion models can again be exploited to essentially inpaint the missing content with new semantically meaningful one.

It is worth noticing that developing an accurate background prediction technology also enables an extra feature in the product: the possibility in some cases to even remove the entire person (face, hair and body) from the picture in the sensitive region. According to the application, such a solution may be appropriate. Case 1: people in an event, full anonymization is desired since the objective there is to show the affluence protecting personal information. Case 2: undesired person in a group of people, complete cancellation of the person in the picture may be preferred. In the picture below, it is shown an example of cancellation with background prediction. Notice that the background prediction technology can be exploited also to produce partial parts surrounding the body.

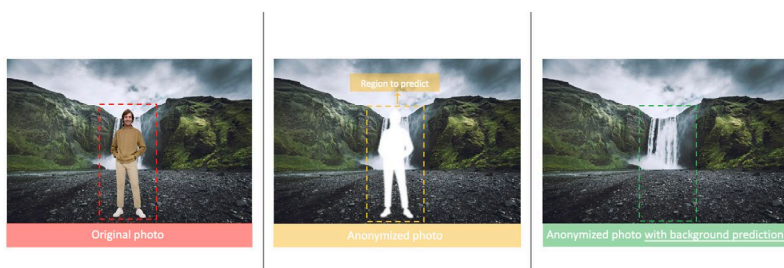


Figure 3: Example of the proposed background prediction with diffusion models applied to cancellation.

Source: own.

References

- Yang, Kaiyu and Yau, Jacqueline and Fei-Fei, Li and Deng, Jia and Russakovsky, Olga (2022). "A Study of Face Obfuscation in ImageNet". *International Conference on Machine Learning (ICML)*
- Zhongzheng Ren and Yong Jae Lee and Micheal S. Ryoo (2018). "Learning to Anonymize Faces for Privacy Preserving Action Detection". *European Conference on Computer Vision (ECCV)*
- Håkon Hukkelås, Rudolf Mester, Frank Lindseth (2019). "DeepPrivacy: A Generative Adversarial Network for Face Anonymization". *International Symposium on Visual Computing (ISVC)*
- Tianxiang Ma, Dongze Li, Wei Wang and Jing Dong (2021). "Face Anonymization by Manipulating Decoupled Identity Representation". *CoRR*
- Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. (2014) "Generative adversarial nets". *Advances in neural information processing systems*. p. 2672–80
- T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen and T. Aila (2020). "Analyzing and Improving the Image Quality of StyleGAN," *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*
- Jonathan Ho and Ajay Jain and Pieter Abbeel (2020). "Denoising Diffusion Probabilistic Models", *Advances in Neural Information Processing Systems*

AGRICULTURAL FIELD DELINEATION USING SATELLITE IMAGERY

MATEJ BATIČ, JAN GERŠAK, MATIČ LUBEJ, ŽIGA LUKŠIČ,
NIKA OMAN KADUNC, DEVIS PERESSUTTI, NEJC VESEL,
SARA VERBIČ

Sinergise, Ljubljana, Slovenia

matej.batic@sinergise.com, jan.gersak@sinergise.com, ziga.luksic@sinergise.com,

matic.lubej@sinergise.com, nika.oman-kadunc@sinergise.com,

devis.peressutti@sinergise.com, nejc.vesel@sinergise.com, sara.verbic@sinergise.com

Abstract Defining the borders of agricultural fields is fundamental for precision agriculture and one of the key parts of the new European Agricultural Policy. The agricultural fields' boundaries are basic building blocks for monitoring agricultural land in the context of climate change, food production and security. The aim of the field delineation process is to automatically determine the borders of agricultural fields from satellite images. It is based on the similarity of spatial, spectral, and temporal properties of pixels belonging to the same field. The basic method was developed within the NIVA project on data from the Sentinel-2 satellite constellation of the European Space Agency. The u-net based deep neural network predicts three image variables from the satellite image: the segmentation of the field, its boundary, and the distance of the segmented image points to the boundary. From these an image of the boundaries of the fields is constructed, either from a single image or from a time series of images. In the post-processing phase, the image prediction is transformed into vector format, which represents the result of the field delineation process.

Keywords:

field delineation,
satellite imagery,
agriculture,
deep learning,
U-Net

DOLOČEVANJE MEJA KMETIJSKIH POLJIN Z UPORABO SATELITSKIH POSNETKOV

MATEJ BATIČ, JAN GERŠAK, MATIC LUBEJ, ŽIGA LUKŠIČ,
NIKA OMAN KADUNC, DEVIS PERESSUTTI, NEJC VESEL,
SARA VERBIČ

Sinergise, Ljubljana, Slovenija
matej.batic@sinergise.com, jan.gersak@sinergise.com, ziga.luksic@sinergise.com,
matic.lubej@sinergise.com, nika.oman-kadunc@sinergise.com,
devis.peressutti@sinergise.com, nejc.vesel@sinergise.com, sara.verbic@sinergise.com

Sinopsis Določitev meja kmetijskih poljin je osnovni proces na področju preciznega kmetijstva ter eden ključnih členov nove Evropske kmetijske politike. Prav tako so meje kmetijskih poljin osnovni gradnik za spremljanje kmetijskih zemljišč v okviru klimatskih sprememb ter prehranske varnosti. Cilj procesa je avtomatska določitev meja kmetijskih poljin iz satelitskih posnetkov. Temelji na podobnosti prostorskih, spektralnih in časovnih lastnostih slikovnih pik, ki pripadajo isti poljini. Osnovno metodo smo razvili v okviru projekta NIVA na podatkih konstelacije satelitov Sentinel-2 Evropske Vesoljske Agencije. Globoka nevronska mreža temelji na u-net arhitekturi in iz satelitskega posnetka napove tri slikovne spremenljivke: segmentacijo poljine, mejo poljine, ter razdaljo segmentiranih slikovnih točk do meje. Iz teh treh napovedi nato sestavimo sliko meja poljin bodisi iz enega posnetka ali pa iz (daljše) časovne vrste posnetkov. V fazi naknadne obdelave slikovno napoved predelamo v vektorski format, ki predstavlja končni rezultat procesa.

Ključne besede:
določevanje meja
kmetijskih poljin,
satelitski posnetki,
kmetijstvo,
globoko učenje,
U-Net

1 Introduction

Defining the borders of agricultural fields is fundamental for precision agriculture and one of the key parts of the new European Agricultural Policy (CAP), which dictates automatic control of agricultural land. The agricultural fields' boundaries are basic building blocks for monitoring agricultural land in the context of climate change, food production and security.

The aim of the field delineation process is to automatically determine the boundaries of agricultural fields from satellite imagery to update existing but outdated datasets of fields, fill in gaps where such data is non-existent, and finally to get a view of how the agricultural landscapes are evolving through time due to anthropogenic activities, climate changes and agricultural practices.

Determination of agricultural fields' boundaries is based on the similarity of spatial, spectral, and temporal properties of pixels belonging to the same field (agricultural land with a single crop). The initial method was developed within the NIVA project¹ on data from the Sentinel-2 satellite constellation of the European Space Agency (ESA). We have since improved the methodology, model, and processing chains. We trained a deep neural network based on AI4Boundaries dataset. The u-net architecture uses satellite imagery to predict three outputs: field segmentation, field boundary, and distance of the segmented pixels to the field border. From these three predictions, we then construct an image of the boundaries of the fields, either from a single image or from a (longer) time series of images. In the post-processing phase, the image prediction is transformed into vector format, which represents the result of the field delineation process.

In the following sections, we will dive-in into a more detailed description of each of these steps and share some of the things we have learned. In 2 Data we will describe the satellite data and ground truth dataset used. The last part of the section will present how we normalize the satellite imagery to facilitate generalizability of the model both through time as well as over larger geographical regions. In 3 Model we will outline the model, its architecture and loss functions. 4 Postprocessing (merging / vectorization) will illustrate the postprocessing of predictions, which allow us to produce results over larger areas (e.g., on continental scale). Lastly,

¹ <https://www.niva4cap.eu>

5 Field delineation as a service will present our field delineation service running on EuroDataCube (EDC), which was used by European Commission Joint Research Centre (JRC) to delineate agricultural fields over the whole Ukraine.

2 Data

2.1 Satellite data

For the source of satellite imagery, we use the openly available Copernicus Sentinel-2 data, accessed through Sentinel Hub services. Sentinel-2 is a land monitoring constellation of two satellites that provide optical imagery with high spatial resolution and high temporal revisit frequency, providing global coverage of the Earth's land surface every 5 days. For delineating fields, we make use of the Level-1C Top of the atmosphere (TOA) reflectance data for all the bands at 10m per pixel resolution (B02, B03, B04, B08).

2.2. Ground truth data

AI4Boundaries (d'Andrimont, 2023), a data set of images and labels readily usable to train and compare field boundary detection models has recently been released by JRC. To train the model, described in next section, we have used the AI4Boundaries ground-truth parcel vectors (2.5 M parcels covering 47105 km²), which have been sourced from openly available Geospatial Aid Application (GSAA) datasets from Austria, Catalonia, France, Luxembourg, the Netherlands, Slovenia, and Sweden for 2019. The data in AI4Boundaries were selected using a stratified random sampling drawn based on two landscape fragmentation metrics, the perimeter/area ratio and the area covered by parcels, thus considering the diversity of the agricultural landscapes across Europe. Training samples of size 256 x 256 were created from Sentinel-2 imagery and ground truth data, as shown in Figure 1.

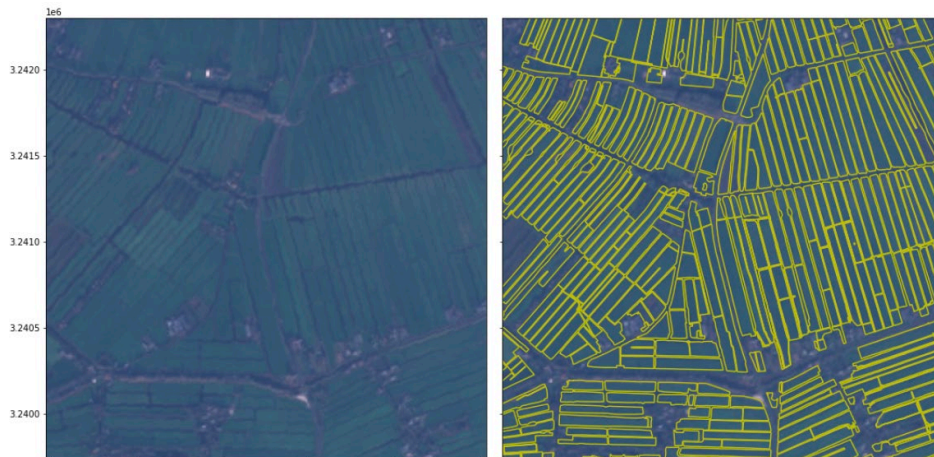


Figure 1: True color Sentinel-2 image (left), overlaid with vector training data from AI4Boundaries (right).

Source: own.

2.3 Normalization

As we want our model to perform well over timestamps taken over the whole year, it is important how the data is normalized. Normalization of the band values can have a significant impact on the network performance and the quality of the field delineation results. The input values for satellite bands are zero-bounded digital numbers and the main characteristics of the band histograms are wide value range, long-tail, and the presence of outlier values. When choosing a normalization method most suited to the properties of the satellite imagery, the aim is to center the distributions and reduce the impact of outliers. In addition, for the normalization procedure to be valid across a wide range of use cases, the training dataset must include imagery from a large geographical region and a long time interval (whole year) to capture both geographical and seasonal variability. We performed an investigation and tested several linear and non-linear normalization schemes on our field delineation model (Oman-Kadunc, 2022). A linear transformation with clipping to 1st and 99th percentile performed best. An example of various normalizations is seen in Figure 2. The observed results of our experiments indicate that mapping the main part of histogram data into the interval $[0, 1]$, but moving outlier values out of this interval (using 1st and 99th percentile) has a large positive effect on the network convergence and performance.



Figure 2: True color Sentinel-2 image before and after various transformations. The last image shows the transformation resulting in best model performance.

Source: own.

3 Model

Reviewing the state-of-the-art in semantic segmentation of temporal images, two main approaches can be considered:

- apply semantic segmentation on each single scene separately and combine predictions temporally at a later stage. A model trained this way should learn to be invariant to the time-period of interest.
- apply semantic segmentation to a temporal stack of images, letting the model extract relevant spatio-temporal features for the task at hand. This approach tends to generate larger and slower models, as the input images contain temporal as well as spatial information (and spectral of course), but implicitly considers temporal dependencies.

The aim of the parcel delineation in CAP practices is generally to monitor agricultural land cover throughout the growing season, but the beginning of the season is of particular interest as it is typically the time when the farmers fill in their applications. A model that can generalize to different time periods seemed therefore useful in this perspective, and that justifies our choice of training a single-scene model and combining temporally the predictions in a subsequent stage. The paper from (Waldner, 2020) represents the state-of-the-art for this approach, and is what we aimed for.

In the initial implementation of the model, we implemented a model architecture as proposed in the above-mentioned paper, which utilizes a u-net backbone (Ronneberger, 2020) with added residual blocks, pyramidal pooling, and conditioned multitasking. While the model performed well on the validation set, we observed

occasional strange behavior of the model when applying it on slightly out-of-distribution data, such as a new region that was not included in the training set. To mitigate such issues, we have decided to revert to a simpler architecture which had only slightly lower scores on the validation set but exhibited a more stable behavior across a variety of real-world datasets.

The architecture that is used in the production model has the u-net backbone with added conditioned multi-tasking outputs as seen on the figure below. Additionally, the max pooling layers within the u-net are replaced with 2D Convolutions with stride 2.

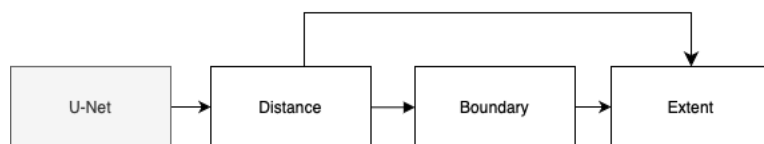


Figure 3: Architecture of the field delineation model. The model outputs three images, showing distance, boundary, and extent of the fields. The loss is computed for each output separately and averaged out.

Source: own.

The model is trained to solve for three conditioned tasks, shown in Figure 3. Its three outputs correspond to the boundaries of the fields, the extent of the fields and the distance from each pixel to the border. The Tanimoto loss, introduced in the ResUNet-a paper (Diakogiannis, 2019), is computed for each of the outputs and averaged to get one loss used for updating the model parameters. During the development we have observed that conditioning of the output had a large positive effect on the quality of the predictions when compared to a version of the model where the outputs were not conditioned.

While the distance is in our case not used when converting the predictions into the final output, it still serves to stabilize the outputs of the model and helps with the training. The model was trained using the Adam optimizer with a fixed learning rate across the duration of training.

The base model is trained to predict images of the same resolution as the input. For example, if we input a $256 \times 256 \times N$ image, the model will return extent, distance and boundary masks of the same width and height. When constructing the training data,

we rasterize the reference vectors to the same spatial resolution as the satellite imagery. We wanted to see if we could extract sub-pixel information by training the model to extract information of a higher resolution than the input. This is done by adding one or multiple pixel shuffle layers to the model architecture and training the model by rasterizing the reference vectors to the target resolution, shown in Figure 4.

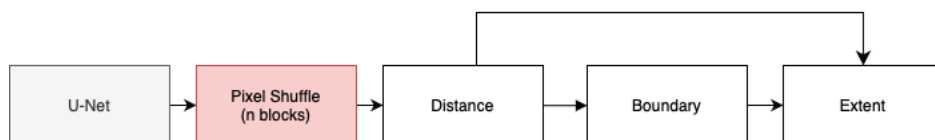


Figure 4: Architecture of the model when we want the model to upscale the output to a higher resolution. Each pixel shuffle block corresponds to a 2x upscaling.

Source: own.

While adding the super-resolution blocks helps the output vectors be more aligned with the actual boundaries, seen in dotted green on Figure 5 below, we also increase the number of parameters and thus the training and prediction time.

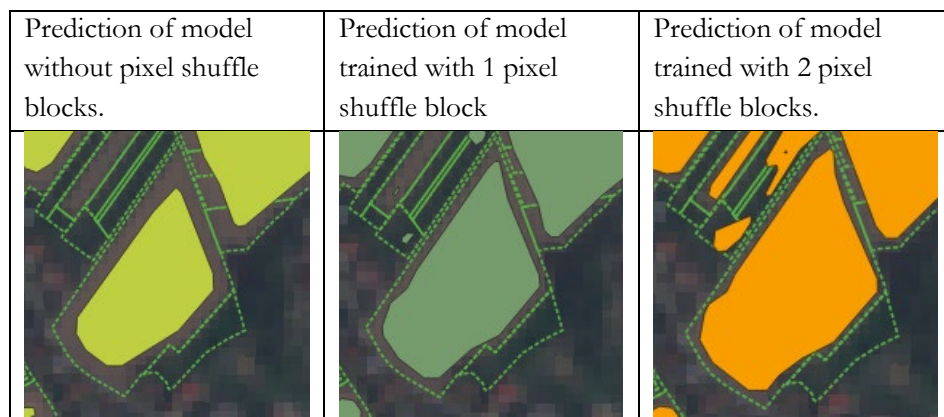


Figure 5: Predictions over the same polygon of a model trained without pixel shuffle blocks (left), model trained with 1 pixel shuffle block (center) and model trained with 2 pixel shuffle blocks (right). We observe that the predicted boundaries are closer to the reference boundaries.

Source: own.

4 Postprocessing (merging / vectorization)

The postprocessing of the single-scene model predictions (i.e., output of the final softmax layer or pseudo-probabilities) is split into two main parts:

- Temporal merging of predictions
- Vectorization of predictions and merging of vectors across EOPatches

4.1 Temporal merging

The model is applied to each available scene during the period of interest. We have observed that each single observation is subject to some degree of noise due to cloud shadows, atmospheric effects or agricultural activity that distorts the real boundaries of the fields. In addition, when running over large areas, it is not possible to choose a single cloudless timestamp that covers the whole area. We tackle these issues by temporally merging the predictions across multiple timestamps. The problem is that the fields themselves are not static through time and can undergo significant changes during only a short period of time, thus the choice of temporal merging method can have a big influence on the results, as can be seen in Figure 6.

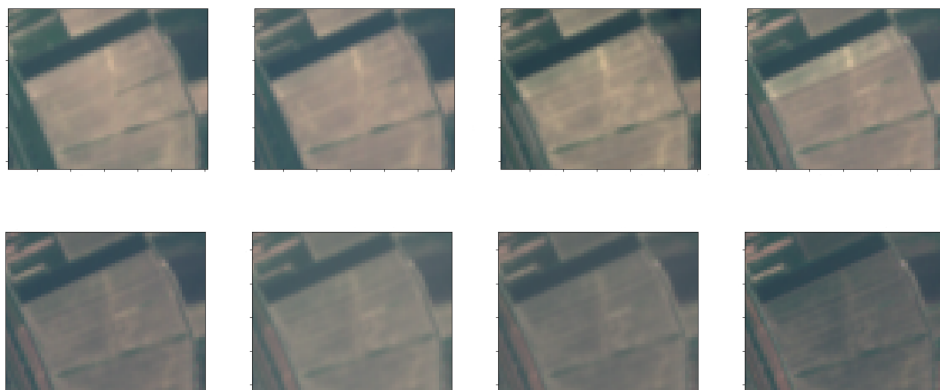


Figure 6: Temporal changes of an agricultural parcel observed for one month (June)

Source: own.

We temporally merge the extent and boundary predictions which we later combine into the final raster mask. The temporal merging is done on pixel level, where the pixels at the same position in each temporal prediction are merged using percentile

statistics. As the optimal temporal merging depends on the use-case at hand, we parametrize the percentile value for both the extent and boundary. For example, for use-cases where the goal is to detect the most representative state over a period, we can choose the 50th percentile (median) for both the extent and boundary. If the goal is to detect the most fields possible (i.e., if a certain field is split in two for only a small amount of time, this split should be detected), we can choose a high percentile value for the boundary and a low percentile value for the extent, as illustrated in Figure 7.

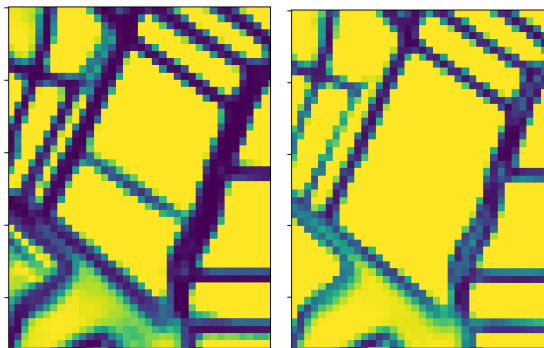


Figure 7: The combined prediction when the percentile value is high for boundary and low for extent (left) and when both percentiles are around 50 (right)

Source: own.

We have also explored alternative merging methods, such as max difference merging which uses the assumption that the timestamps where the difference between extent and boundary are the highest are the ones where the model is the most confidently distinguishing between the extent and the boundary. So given a temporal stack, you take the extent and boundary at the position where there is the biggest difference between the two.

Another method that we developed is merging with a rolling window, which is designed with the goal of detecting all possible stable boundaries within a period. The idea behind this method is to perform temporal merging by utilizing a rolling temporal window to smooth out outliers and to choose a stable period. The procedure computes the mean and standard deviation of extent and boundary for values inside each of the temporal windows. The window providing the best estimate

is chosen to be the one with a high boundary value, low extent value and a low standard deviation (representing a stable state).

4.2 Vectorization

From the step above, given a time interval, we can aggregate predictions and obtain a single pseudo-probability image for extent, boundary, and distance. We now combine these and obtain a vector layer for the entire country. To obtain smoother vectors, the pseudo-probabilities are combined into a single image as

$$p = 1 + p_{\text{extent}} - p_{\text{boundary}}, \quad (1)$$

as we didn't use the distance masks in this iteration. This resulting image has continuous values in the $(0, 2)$ range, and can be treated as a level set functional that can be sectioned to obtain nice and smooth contours. To obtain the contours from the raster image, we used the GDAL `contour2` utility, using parameters that gave best overlaps with the GSAA vectors.

Another very useful feature of GDAL we used is the Virtual Raster Format (VRT), which allowed us to build a virtual raster containing the merged functionals of all EOPatches. This way the predictions could be blended into a smooth functional even at the borders of EOPatches. Using VRT we can run contouring parallelizing over smaller and overlapping areas, generating vector shapes that are matching over the overlapping area. To obtain a single vector layer, the overlapping geometries were merged performing a geometrical union.

5 Field delineation as a service

Lastly, we have put all the pieces together using `eo-grow`, Earth observation framework for scaled-up processing in Python³. In a nutshell, the following steps to produce boundaries are performed:

² https://gdal.org/programs/gdal_contour.html

³ <http://github.com/sentinel-hub/eo-grow/>

- split the area of interest (AOI) into a regular grid to speed up processing through massive parallelization;
- download remote sensing imagery for the time interval of interest using Sentinel Hub batch processing API that outputs the imagery directly to our AWS S3 bucket;
- predict and post-process agricultural parcel boundaries on remote sensing imagery for the time interval of interest, parallelized over the EOPatches in the grid;
- perform vector merging over the whole area into a single result.

eo-grow splits the AOI into a regular grid of EOPatches, like shown in Figure 8 below. Sentinel Hub batch processing API delivers available satellite imagery for each EOPatch into directly into an AWS S3 bucket. In the next step the data is fed to the model to produce predictions, which are then post-processed and temporally merged. Vectorization is performed, and the results finally merged into a single file that can be used in GIS software.

An algorithm using the approach above is available as a service on EuroDataCube⁴. It was used within the EO4UA initiative to delineate agricultural fields over Ukraine for years 2016-2022. The web application showing results can be seen in Figure 9. The dataset facilitated further research into how war is affecting the agricultural landscape in Ukraine, their local food producing capabilities and, consequently, global food supplies.

⁴ <https://collections.eurodatacube.com/field-delineation/>

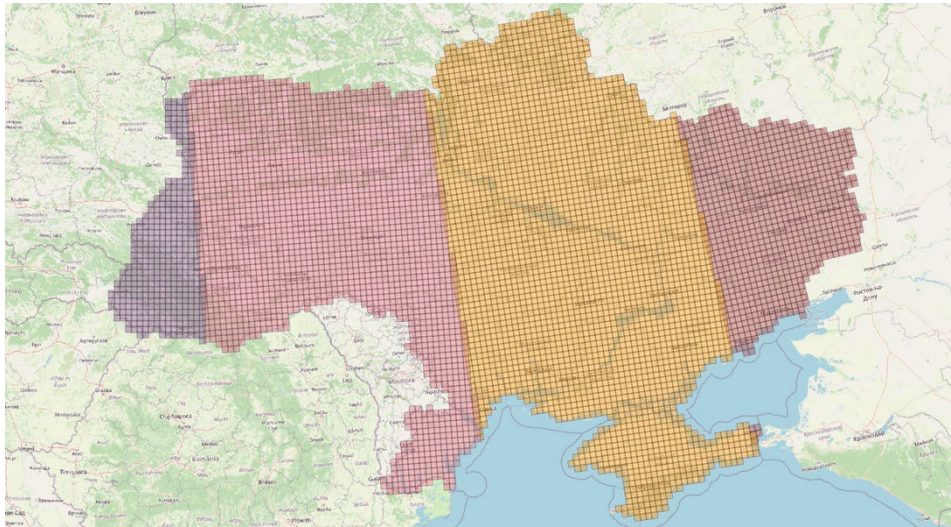


Figure 8: Ukraine split into tiles of 10km x 10km, tiles are in their own UTM zone.

Source: own.

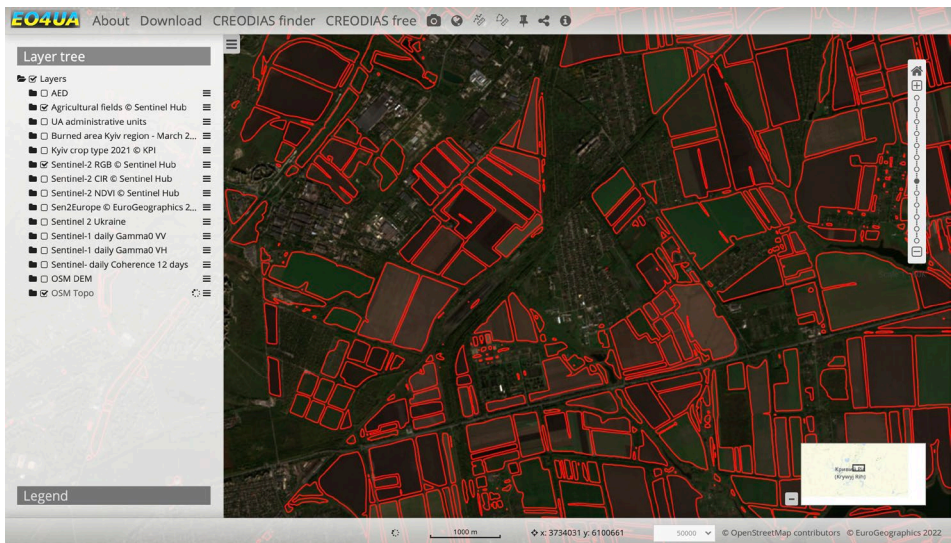


Figure 9: Web viewer of the EO4UA initiative⁵, showing Sentinel-2 true color imagery overlaid with delineated agricultural fields.

Source: own.

⁵ <https://www.co4ua.org/>

References

- d'Andrimond, R. (2023). AI4Boundaries: an open AI-ready dataset to map field boundaries with Sentinel-2 and aerial photography. *Earth Syst. Sci. Data*, 15, 317–329. doi:10.5194/essd-15-317-2023
- Diakogiannis, F. (2020). ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 162, 94-114. doi:10.1016/j.isprsjprs.2020.01.013
- Oman-Kadunc, N. (2022). How To Normalize Satellite Images for Deep Learning [blog post]. Retrieved from <https://medium.com/sentinel-hub/how-to-normalize-satellite-images-for-deep-learning-d5b668c885af>
- Waldner, F. (2020). Deep learning on edge: Extracting field boundaries from satellite images with a convolutional neural network. *Remote sensing of environment*, 245. doi:10.5194/essd-15-317-2023
- Ronneberger, O. (2020). U-net: Convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds) *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. MICCAI 2015. *Lecture Notes in Computer Science*, vol 9351. Springer, Cham. doi:10.1007/978-3-319-24574-4_28

DOPOLDANSKA IN POPOLDANSKA SEKCIJA

Industrijske aplikacije

Medicinske in biomedicinske aplikacije

Drugo

Študentske aplikacije

MODELIRANJE DINAMIKE PEČENJA ZA DOLOČANJE STANJA PEČENJA Z MREŽAMI LSTM

NINA SEDEJ,^{1,2} GREGOR KOPOREC,² JANEZ PERŠ¹

¹ Univerza v Ljubljani, Fakulteta za elektrotehniko, Ljubljana, Slovenija
nina.sedej@gorenje.com, janez.pers@fe.uni-lj.si

² Gorenje d.o.o, Ljubljana, Slovenija
nina.sedej@gorenje.com, gregor.koporec@gorenje.com

Sinopsis Osnovni način, da dosežemo dobre rezultate pri pečenju je prilagoditev časa – če želi ljubiteljski kuhar bolj zapečene piškote bo podaljšal čas peke. Ta pristop ne zagotavlja vedno istih rezultatov in lahko vodi v preveč ali premalo zapečene jedi. Za reševanje tega problema je bilo razvitih že več sistemov računalniškega vida, ni pa še bilo izvedene sistematične študije, ki bi razviti sistem primerjala z izkušenim domačim kuharjem. V tem delu predstavimo sistem računalniškega vida, ki je sestavljen iz pečiče s kamero, sistema za zajemanje slik in globokih nevronskega modelov. Delovanje sistema primerjamo z modelom ljubiteljskega kuharja. Ker se videz jedi v pečiči spreminja skozi čas, poleg konvolucijskega modela CNN uporabimo dve vrsti modelov, ki na vходу sprejmeta zaporedje slik - CNN-LSTM in ConvLSTM. Rezultati kažejo, da model ConvLSTM prekaša model ljubiteljskega kuharja za 5 odstotnih točk v metriki F1. Da so modeli primerni za spremljanje kvalitete jedi v pečiči, morajo imeti sposobnost učenja dinamike pečenja.

Ključne besede:
dinamika pečenja,
stanje pečenja,
CNN,
CNN-LSTM,
ConvLSTM,
razvrščanje

MODELING BAKING DYNAMICS FOR DETERMINING BAKING STATE USING LSTM NETWORKS

NINA SEDEJ,^{1,2} GREGOR KOPOREC,² JANEZ PERŠ¹

¹ University of Ljubljana, Faculty of Electrical Engineering, Ljubljana, Slovenia
nina.sedej@gorenje.com, , janez.pers@fe.uni-lj.si

² Gorenje d.o.o, Ljubljana, Slovenia
nina.sedej@gorenje.com, gregor.koporec@gorenje.com

Abstract The conventional approach of getting a satisfactory baking result is to adapt baking time. If the home cook prefers the cookies to be darker, they will prolong the baking time. This approach doesn't guarantee the same results every time and can result in an underbaked or burnt dish. To solve this issue computer vision-based systems have already been developed, but there isn't any systematic study that would compare a developed computer vision system to an experienced home cook. We present a computer vision system consisting of an oven with a camera, an image acquisition system and deep learning models. We compare it to a home cook model. To model baking dynamics we also use models, that take a sequence of images as their input: CNN-LSTM and ConvLSTM. We conclude that ConvLSTM outperforms the home cook model by 5 percentage points in F1 metric. A crucial attribute for machine learning models to be successfully used for monitoring baking quality is learning baking dynamics.

Keywords:

baking dynamics,
baking state,
CNN,
CNN-LSTM,
ConvLSTM,
classification

1 Uvod

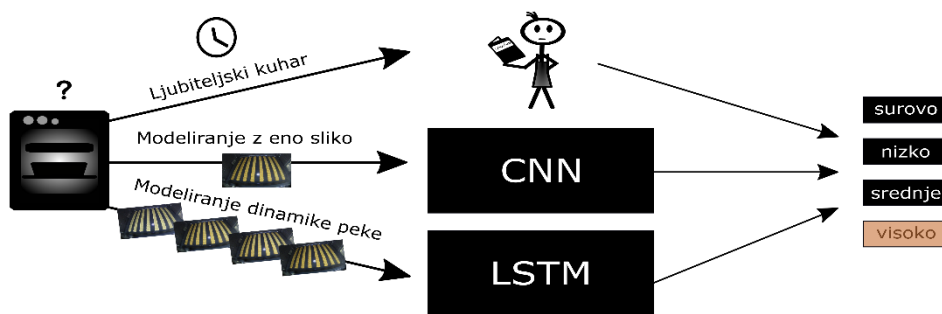
Peka jedi je kompleksen nelinearen proces (Pylar in Gorton, 2008), ki ga lahko v splošnem razdelimo na dve fazi. V prvi fazi je dinamika procesa počasna, saj ne opazimo vidnih sprememb. Sledi faza hitre dinamike, kjer zaradi kompleksnih kemičnih procesov živilo poleg okusa spremeni tudi svoj videz (obliko, barvo in teksturo) (Purlis, 2010). Videz tako postane bistven dejavnik za modeliranje stanja pečenja (Purlis, 2010). Izraz stanje pečenja pomeni oceno gotovosti jedi (surova, nizko, srednje ali visoko zapečena).

Ugotavljanje stanja pečenja v dinamični fazi peke je pomemben vidik poklicnega kuharja, saj se poklicni kuhar lahko nauči, kako čas pečenja vpliva na kakovost gotove jedi. Čas pečenja tako postane bistven parameter recepta, s katerim poenostavimo zapletenost postopka peke. Recept lahko uporabi ljubiteljski kuhar, pri čemer lahko s spremembo časa pečenja jed prilagodi svojim željam. Na primer, če ima rajši temnejše piškote, jih lahko peče nekaj minut dlje, kot piše v receptu.

Prilagoditev časa peke za zagotavljanje zelenih rezultatov ima svoje pomanjkljivosti: (i) Zgolj sledenje receptu nam ne zagotavlja enakih rezultatov, saj je postopek peke odvisen tudi od drugih dejavnikov (npr. vrsta pečice, temperatura, vrsta pladnja, variacija v sestavinah v različnih državah, razlika v masi sestavin, ipd.). Jed, ki je pripravljena po enakem receptu, je zaradi tovrstnih dejavnikov lahko gotova nekaj minut prej ali pozneje. (ii) Dinamična faza lahko ob koncu pečenja povzroči hitre spremembe videza jedi. Ljubiteljski kuharji tako lahko hitro podcenijo dinamiko pečenja, rezultat pa je preveč pečena ali celo zažgana jed. V tem primeru jed ni samo neužitna, ampak tudi nevarna za zdravje (Rudén, 2004). (iii) Ko kuharji odpirajo pečico, da ugotovijo, ali je jed gotova, povzročijo padec temperature, kar zmoti proces pečenja.

Za reševanje tovrstnih problemov so bile zato razvite podporne tehnologije na osnovi računalniškega vida. Običajno delujejo tako, da sistem računalniškega vida zajame vizualno informacijo, iz katere s tehnikami stojnega učenja razberemo spremembe stanja jedi. Vendar pa do sedaj še ni bilo izvedene študije, kjer bi sistematsko primerjali razviti sistem računalniškega vida, z metodo prilagajanja časa, ki jo uporabljajo ljubiteljski kuharji. Tako ni jasno, *ali je sistem računalniškega vida sploh primerna rešitev za določanje stanja jedi v pečici?*

Da bi odgovorili na to vprašanje, v tem delu sistematično primerjamo trenutno najboljši sistem računalniškega vida z metodo prilagajanja časa pečenja, ki ga uporabljajo domači kuharji – *model ljubiteljskega kuharja*. Slika 1 prikazuje shematični prikaz prikazanega pristopa. V ta namen smo razvili sistem računalniškega vida, ki je sestavljen iz pečice s kamero in prilagojeno osvetlitvijo ter modela globoke nevronske mreže za razpoznavanje stanja pečenja: je jed surova ali nizko, srednje ali visoko zapečena. Model ljubiteljskega kuharja simulira odločitve izkušenega ljubiteljskega kuharja in predstavlja zahtevno osnovo za primerjalno analizo modelov globokih nevronske mreže.



Slika 1: Shematični prikaz predlaganega pristopa. Primerjamo tri modele za spremljanje stanja pečenja: ljubiteljskega kuharja, ki peče po receptu, CNN, ki procesira po eno sliko naenkrat in dve LSTM arhitekturi, ki procesirata časovno sosledje več slik.

Vir: lasten.

Ker se poklicni kuharji učijo kompleksne dinamike pečenja z opazovanjem pečenja skozi čas, pričakujemo, da je mogoče učinkovito modelirati dinamiko pečenja z modeli, ki izkoriščajo časovno sosledje podatkov. Domnevamo, da bodo arhitekture LSTM, ki na vnosu sprejmejo časovno zaporedje slik, izboljšale rezultate modelov CNN, ki delujejo na posamičnih slikah. Da bi preverili to predpostavko, v tem delu analiziramo dve različni arhitekturi LSTM: CNN-LSTM in ConvLSTM.

Rezultati študije kažejo, da modeli CNN niso primerni za razpoznavanje stanja jedi, saj so slabši od modela ljubiteljskega kuharja. Arhitekture ConvLSTM so po drugi strani obetavne, saj ljubiteljskega kuharja prekašajo za 5 odstotnih točk v metriki F1. Rezultati potrjujejo domnevo: *Ključna značilnost modelov za spremljanje stanja peke je sposobnost učenja dinamike pečenja.*

2 Sorodna dela

Kakovost izdelkov, ki se pečejo v pečici lahko določamo iz meritev različnih senzorjev, kot so temperaturna sonda, spektrometer ali kolorimeter (Ureta et al., 2014; Ureta et al., 2016). Težava pri tovrstnih analizah je, da so destruktivne. Kadar želimo določiti lastnosti jedi, jo moramo vzeti iz pečice, to pa vpliva na proces pečenja. V izogib težavam kontaktnega merjenja, je bilo predlaganih več neinvazivnih metod računalniškega vida (Mogol in Gökmen, 2014).

Sistem računalniškega vida lahko uporablja vizualne lastnosti kot sta barva in oblika hrane za oceno njenih lastnosti. Shafiee et al. (2014) so uporabili sistem računalniškega vida za določanje barve medu in uporabili preprost model nevronske mreže za napovedovanje lastnosti medu in določitev njegove kakovosti. Purlis in Salvadori (2009) sta predlagala kolorimetrično metodo za modeliranje in napovedovanje sprememb v barvi kruhove skorje med peko s pomočjo matematičnega modela za kinetično rjavenje. Paquet-Durand et al. (2012) so pri spremljanju procesa peke žemljic uporabili tudi informacijo o spremembi velikosti.

Konvolucijske nevronske mreže (angl. Convolutional Neural Networks – CNN) so zelo uspešne pri reševanju številnih nalog, povezanih s slikami (Krizhevsky et al., 2012; He et al., 2016). Cotrim et al. (2020) so zato predlagali kolorimetrični pristop, v katerem s pomočjo CNN razvrščajo barvo kruhove skorje v sedem različnih stanj. Kljub temu do sedaj še ni bilo izvedene sistematične študije, kjer bi primerjali modele globokih nevronskih mrež s sposobnostjo ljubiteljskega kuharja, da s prilagoditvijo časa pečenja dobi optimalne rezultate.

Razredi modelov, ki so pokazali dobro delovanje s časovno soslednimi podatki, so mreže z dolgim kratkoročnim spominom (angl. Long Short-Term Memory - LSTM) (Hochreiter in Schmidhuber, 1997). Za uporabo mrež LSTM z zaporedji slik je bilo predlaganih več rešitev, med katerimi izstopata različici CNN-LSTM in ConvLSTM. Različica CNN-LSTM je sestavljena iz komponente CNN, ki jo uporabimo kot določevalnik značilk, in LSTM komponente, od koder dobimo končne rezultate razpoznavalnika (Donahue et al., 2017). Različica ConvLSTM je arhitekturno bolj podobna običajni LSTM celici, le da uteži nadomestijo konvolucijska jedra, kar omogoča pretok podatkov skozi časovno mrežo v izvorni 3D obliki (Shi et al., 2015).

Arhitekture CNN-LSTM so bile uporabljene pri različnih opravilih, kot je opisovanje (Gao et al., 2017; Donahue et al., 2017) in razvrščanje video posnetkov (Ng et al., 2015). Po drugi strani so (Yadav et al. 2022) model ConvLSTM uporabili za razpoznavanje človeških kretenj na podlagi posnetkov vzorcev gibanja. (Shi et al. 2015) so model ConvLSTM uporabili za napovedovanje padavin na podlagi prostorsko-časovnih radarskih podatkov. V okviru obeh raziskav so ugotovili, da je model ConvLSTM dosegel boljše rezultate kot model CNN-LSTM. Nismo zasledili raziskav, v katerih bi bila katera od različic mreže LSTM uporabljena za spremljanje kakovosti hrane pri pečenju v pečici.

3 Podatkovna zbirka

3.1 Sistem za zajem podatkov

Sistem za zajem podatkov je bil sestavljen iz prototipa pečice, digitalne kamere in računalnika. Prototip pečice s kamero je prispevalo podjetje Gorenje d.o.o. Kamera je bila nameščena v zgornji del ohišja pečice in je zajemala slike skozi odprtino, ki je bila izolirana s steklom, odpornim na visoke temperature. Za izboljšanje kakovosti slike smo obstoječo osvetlitev v pečici dopolnili z dvema LED trakovima. Uporabili smo 1/2,7-palčno RGB kamero, ki zagotavlja ločljivost 1920×1080 slikovnih pik. S kamero smo vsakih 30 sekund zajemali fotografije plitkega emajliranega pekača dimenzije 460×360 mm, na katerem se je pekla jed. Ker je delovna temperatura kamere 105 °C, temperature v pečici pa se lahko gibljejo tudi do 400 °C (piroliza), smo za hlajenje kamere izdelali ventilatorski sistem. Kamera je bila povezana na računalnik Raspberry Pi 4, ki je nadzoroval zajemanje slik s programsko opremo Motion (motion-project, 2020) različice 4.3.2.

3.2 Peka piškotov

Pekli smo brizgane piškote, ki so bili pripravljene po standardnem receptu za določanje enakomernosti porazdelitve toplote na bremenu. Izraz pečenje označuje celoten postopek peke od postavitve pladnja s surovo jedjo v pečico do trenutka, ko jed doseže zlatorjavo bravo (kategorija visoko zapečeno) in jo vzamemo iz pečice. Posamezna peka je trajala okvirno 33 min. S posamezno peko smo pridobili v povprečju 60 slik. Da bi bili pridobljeni podatki čim bolj raznovrstni, smo pekli piškote na 35 različnih načinov. Ti vključujejo tri različne programe peke (predgretje,

klasična peka in vroči zrak), dve različni temperaturi (140 °C in 145 °C), pekača dveh različnih barv (siva in črna) ter dva nivoja v pečici (srednji in spodnji nivo). Poleg tega smo uporabili dve različni razporeditvi piškotov na pekaču (pravokotna in vzporedna) ter različno količino piškotov (od 5 do 11 vrstic).

Na koncu vsakega poskusa smo piškote prestavili iz pečice v testno komoro, kjer smo s spektrofotometrom izmerili njihovo barvo in določili končno stanje jedi. Na podlagi meritve končnega stanja smo nato določili kategorije slik (glej razdelek 3.4 za podrobnosti).

3.4 Podatkovna zbirka OvenRecognition

Podatkovna zbirka OvenRecognition je sestavljena iz 32990 slik brizganih piškotov. Primeri vzorcev iz podatkovne zbirke so prikazani na sliki 2. 28230 slik je namenjenih učenju modelov, 4760 slik sodi v testno množico. Vsak vzorec vsebuje podatke o času in programu peke, ID oznako pečenja ter stanje zapečenosti. Stanje zapečenosti določajo štiri kategorije: surovo, nizko zapečeno, srednje zapečeno in visoko zapečeno. V nasprotju s kategorijo surovo, ostale kategorije predstavljajo užitne vzorce. Podatkovna zbirka je glede na stanje zapečenosti močno neuravnotežena (npr. 22970 vzorcev spada v kategorijo surovo), kar je posledica dinamike procesa pečenja.



Slika 2: Primeri slik iz zbirke podatkov OvenRecognition, ki predstavljajo vzorce iz kategorij surovo, nizko zapečeno, srednje zapečeno in visoko zapečeno.

Vir: lasten.

4 Modeliranje procesa pečenja

4.1 Model ljubiteljskega kuharja

Kot osnovo za primerjavo modelov smo uporabili model ljubiteljskega kuharja, ki razvršča slike na podlagi časa, ki je preteklo od začetka pečenja. Namen te metode je simulirati uporabnika pečice, ki upošteva čas peke, kot je naveden v receptu. Časovne

meje med kategorijami smo izbrali na način, da zmanjšamo število napačno razvrščenih slik na učni množici podatkovne zbirke. Na ta način smo simulirali proces odločanja izkušenega ljubiteljskega kuharja, in jih hkrati naredili bolj nepristranske. Meje smo postavili na naslednje vrednosti: 0–25 min za surovo, 25,5–28 min za nizko zapečeno, 28,5–33 min za srednje zapečeno ter 33,5 min in več za visoko zapečeno kategorijo.

4.2 CNN

Za model CNN smo izbrali EfficientNet_b4, ki je bil prednaučen na podatkovni zbirki ImageNet (Deng et al., 2009). Model prihaja iz družine razširljivih modelov, ki z relativno malo hiperparametri dosegajo dobre rezultate (Tan in Le, 2019). Zadnji, polno povezan sloj smo doučili na podatkovni zbirki OvenRecognition. Začetna učna stopnja je bila 10^{-2} . Učno stopnjo smo zmanjšali za faktor 10 vsakič, ko se izguba ni izboljšala v 20 zaporednih epohah učenja. Učne podatke smo na začetku vsake epohe premešali. Neuravnoteženost podatkovne zbirke smo rešili z uteževanjem kriterijske funkcije navzkrižne entropijske izgube (angl. Cross Entropy Loss). Pri tem smo uporabili inverzno relativno frekvenco razreda.

4.3 CNN-LSTM

Uporabili smo dva različna modela CNN-LSTM. V prvem modelu, poimenovanem AlexCNN-LSTM smo za luščenje značilik uporabili model AlexNet (Krizhevsky et al., 2012). V drugem modelu, EfficientCNN-LSTM, smo za luščenje značilik uporabili model EfficientNet_b4 (Tan in Le, 2019). Obe CNN komponenti sta bili prednaučeni na podatkovni zbirki ImageNet (Deng et al., 2009). Za vhodno zaporedje slik smo uporabljali 4 časovno sosledne slike. LSTM komponente modelov smo doučili na podatkovni zbirki OvenRecognition. Doučevanje je potekalo na enak način kot pri modelu CNN, le da je bila začetna učna stopnja 10^{-3} .

4.4 ConvLSTM

Uporabili smo model ConvLSTM iz (Shi et al., 2015), ki smo ga prilagodili tako, da je sestavljen iz začetnega konvolucijskega sloja, ki mu sledi sloj ConvLSTM. Naučili smo dva modela ConvLSTM z različno velikostjo vhodnega zaporedja slik. ConvLSTM_n2 predstavlja model, ki uporablja 2 časovno sosledni sliki za razpoznavanje. Z ConvLSTM_n4 smo označili model, ki uporablja 4 časovno

sosledne slike. Modela ConvLSTM smo naučili izključno na podatkovni zbirki OvenRecognition. Učenje je potekalo na enak način kot pri modelu CNN, le da je bila začetna učna stopnja 10^{-4} .

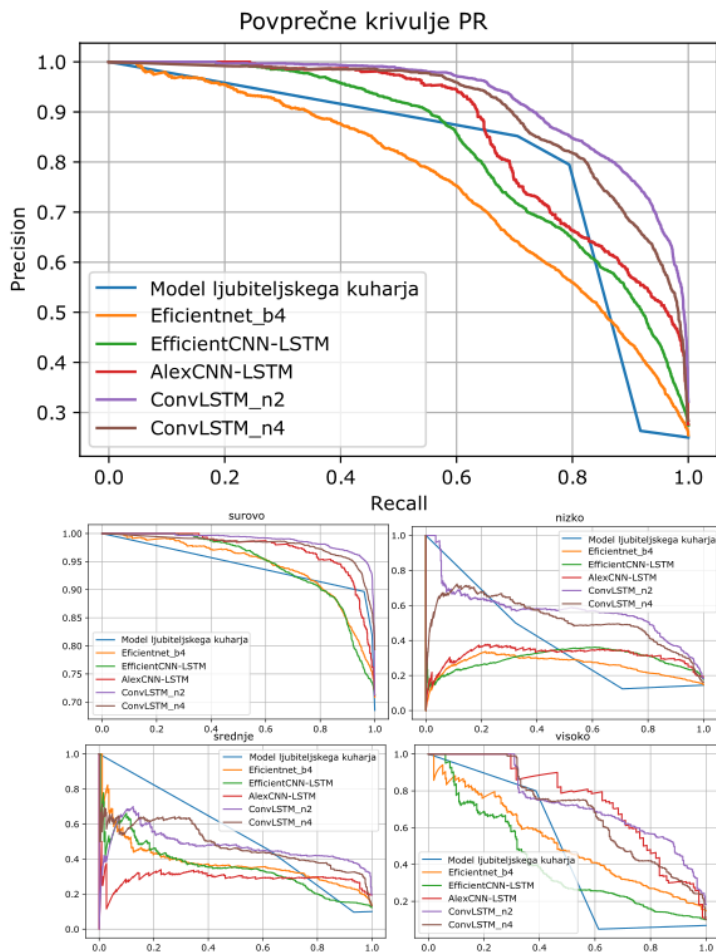
5 Eksperimenti

Da bi ocenili, ali je uporaba globokih nevronskih mrež primerna rešitev za določevanje stanja jedi v pečici, smo najprej primerjali izbran model CNN z modelom ljubiteljskega kuharja. Pričakovali smo slabe rezultate, saj menimo, da modeli CNN po naravi ne zmorejo natančneje modelirati dinamike pečenja, ki jo lahko zasledimo z opazovanjem skozi čas. Po drugi strani smo pričakovali, da se dinamike pečenja lahko naučijo modeli CNN-LSTM, saj razpoznavanje na podlagi časovnega zaporedja slik. Pri tovrstnih modelih smo dodatno preverili, kako različna kapaciteta določevalcev značilnik vpliva na rezultate. Ker sorodna dela nakazujejo na dodatne izboljšave pri uporabi modelov ConvLSTM, smo preverili tudi njihovo delovanje pri razpoznavanju stanja pečenja. Pri eksperimentih s tovrstnimi modeli smo preverili še vpliv dolžine časovnega zaporedja slik.

Za evaluacijo modelov smo uporabili pet različnih metrik: Precision, Recall, F1, krivuljo PR in normalizirano matriko zamenjav (angl. confusion matrix).

6 Rezultati in diskusija

Slika 3 prikazuje povprečne krivulje PR in krivulje PR za vsak razred posebej. Rezultati povprečnih krivulj PR nakazujejo, da model Efficientnet_b4 ni zmožen dosegati rezultatov ljubiteljskega kuharja. S podrobno analizo krivulj PR za posamezne razrede smo ugotovili, da Efficientnet_b4 za surovo jed prekaša ljubiteljskega kuharja do vrednosti Recall 0,8. Menimo, da bi to lahko bila posledica neuravnoteže podatkovne zbirke, saj je surova kategorija močno zastopana. Delovanje modela Efficientnet_b4 za ostale kategorije je bistveno slabše od ljubiteljskega kuharja. Rezultati, prikazani v tabeli 1, prav tako nakazujejo na slabo delovanje modela Efficientnet_b4. Razlika med modelom Efficientnet_b4 in ljubiteljskim kuharjem v metriki F1 je 11,34 odstotnih točk. To potrjuje naša pričakovanja in nakazuje, da je učenje dinamike peke ključna značilnost za uspešno spremljanje stanja pečenja.



Slika 3: Krivulje PR za surovo, nizko, srednje in visoko zapečeno kategorijo. Večja površina pod krivuljo v splošnem pomeni boljše delovanje. Modela ConvLSTM dosegata boljše rezultate za večino vrednosti Recall-a. Modeli CNN in LSTM ne dosežajo boljših rezultatov od modela ljubiteljskega kuharja.

Vir: lasten.

Tabela 1: Vrednosti metrik Precision (P), Recall (R) in F1 za posamezne modele. Modela ConvLSTM dosežeta najboljše rezultate v vseh metrikah.

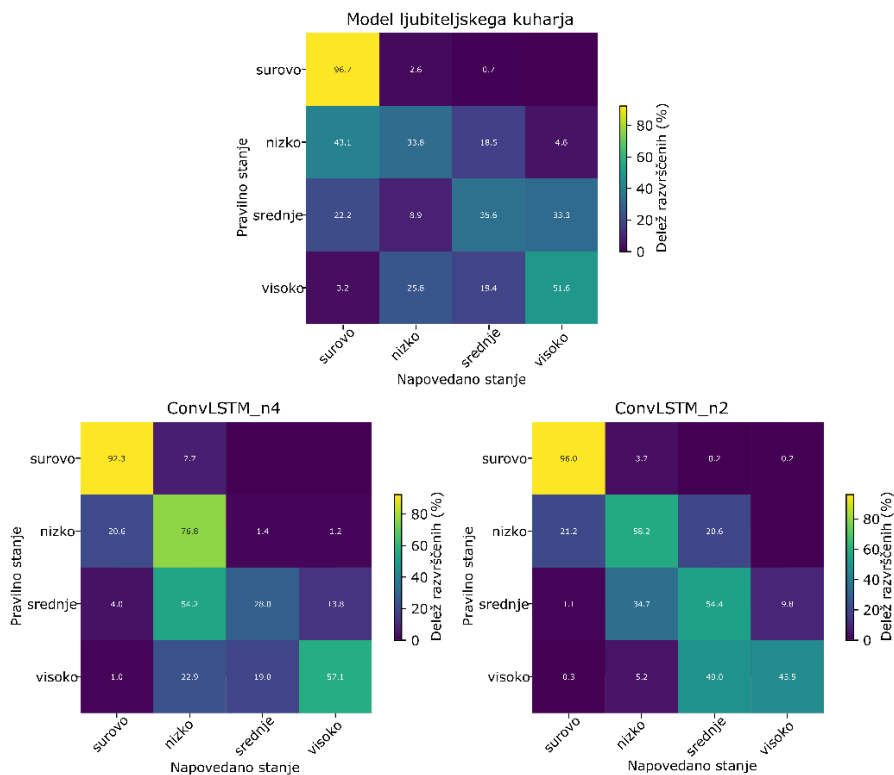
Model	P	R	F1
Model ljubiteljskega kuharja	66,21	57,33	61,45
EfficientNet_b4	47,73	52,73	50,11
AlexCNN-LSTM	63,27	39,16	48,37
EfficientCNN-LSTM	55,61	41,53	47,56
ConvLSTM_n2	68,07	63,52	65,72
ConvLSTM_n4	69,72	63,54	66,49

vir: lasten

Pri analizi rezultatov modelov CNN-LSTM smo ugotovili, da tovrstni modeli prekašajo model CNN EfficientNet_b4 po krivuljah PR in metriki Precision, metriki Recall in F1 pa nakazujeta slabše delovanje. Opazimo lahko, da je natančnost modelov CNN-LSTM slabša od ljubiteljskega kuharja na območju Recall od 0,6 do 0,8. Pri primerjavi med različnimi kapacitetami določevalnikov značilk v modelih CNN-LSTM smo ugotovili, da povečana kapaciteta daje slabše rezultate. Rezultati bi lahko nakazovali, da se CNN-LSTM modeli do določene mere učijo dinamike pečenja, vendar ne dovolj. Verjetno ima velik vpliv komponent CNN modelov, ki ne zmore luščiti značilk z vsebovano dinamiko pečenja. Ločevanje značilk na posamezne razrede v nadaljnjih slojih modela pa je posledično oteženo. Vpliv komponent CNN se lahko veča z večanjem kapacitete določevalnika, kar bi bil lahko vzrok slabšega delovanja modela EfficientCNN-LSTM.

Po drugi strani modeli ConvLSTM presegajo rezultate modela ljubiteljskega kuharja, modelov CNN-LSTM in modela CNN EfficientNet_b4 po krivuljah PR. Boljše delovanje nakazujejo tudi ostale metrike. Rezultati med ConvLSTM_n4 in modelom ljubiteljskega kuharja se razlikujejo za 3,51 odstotne točke v metriki Precision, 6,21 odstotnih točk v metriki Recall in 5,01 odstotnih točk v metriki F1. Za natančnejšo primerjavo med modelom ljubiteljskega kuharja in modeli ConvLSTM smo primerjali njihove matrike zamenjav, ki so prikazane na sliki 4. Matrike so izraziteje diagonalne za modele ConvLSTM, kar pomeni, da je bilo pravilno razvrščenih več vzorcev. Še več, modeli ConvLSTM razvrstijo bistveno manj užitnih vzorcev (nizko, srednje in visoko zapečeno) med surove kot model ljubiteljskega kuharja.

Da bi pridobili dodaten vpogled v delovanje modela ConvLSTM smo preverili še vpliv dolžine časovnega zaporedja slik. Model ConvLSTM_n4 je dosegel boljše rezultate kot ConvLSTM_n2 v vseh metrikah (tabela 1), vendar razlike niso statistično značilne. Sklepamo lahko, da zmanjšanje število vhodnih slik ni bistveno vplivalo na sposobnost modeliranja dinamike pečenja modela ConvLSTM. Iz slike 4 je razvidno, da je ConvLSTM_n4 boljši pri razvrščanju kategorije nizke in visoke zapečenosti, vendar pa označi več surovih vzorcev kot užitne, kar pa ni zaželeno. Model s krajšim časovnim zaporedjem slik zato predstavlja boljšo izbiro za spremljanje stanja jedi med pečenjem v pečici.



Slika 4: Primerjava matrik zamenjav za model ljubiteljskega kuharja in modela ConvLSTM_n4 in ConvLSTM_n2.

Vir: lasten.

7 Zaključek

Osnovni pristop za doseganje zadovoljivega rezultata pečenja v pečici je prilagajanje časa pečenja. Ta pristop ne zagotavlja vedno enakih rezultatov in nemalokrat povzroči zažgano jed. Da bi rešili to težavo smo razvili sistem računalniškega vida s katerim brezkontaktno zajemamo slike jedi v notranjosti pečice. S pomočjo modela globokega učenja razpoznavamo stanje jedi v pečici. Razviti sistem računalniškega vida sistematično primerjamo z modelom ljubiteljskega kuharja, ki simulira odločitve izkušenega uporabnika pečice.

Poklicni kuharji se učijo kompleksne dinamike učenja z opazovanjem pečenja jedi skozi čas, zato pričakujemo, da bodo arhitekture LSTM, ki izkoriščajo časovno sosledje podatkov, izboljšale rezultate modela CNN. Da bi preizkusili to hipotezo

smo poleg modela CNN naučili in analizirali tudi štiri modele arhitektur CNN-LSTM in ConvLSTM. Rezultati študije kažejo, da modeli CNN in CNN-LSTM niso primerni za razpoznavanje stanja pečenja. Arhitekture ConvLSTM so po drugi strani dosegle obetavne rezultate, saj med drugim v metriki F1 prekašajo model izkušenega ljubiteljskega kuharja za 5 odstotnih točk. Končna ugotovitev te raziskave je, da je učenje dinamike peke ključna značilnost za uspešno uporabo modelov strojnega učenja za spremljanje stanja jedi med pečenjem v pečici.

Ta ugotovitev odpira nove raziskovalne priložnosti, s katerimi bi lahko dodatno izboljšali spremljanje kakovosti peke. Ker se modeli LSTM osredotočajo le na nekaj zaporednih slik, bi lahko razvili arhitekture, ki bi spremljale celoten proces pečenja.

Opombe

Raziskavo je omogočilo podjetje Gorenje d.o.o.

Literatura

- Cotrim, W. d., Minim, V. P., Felix, L. B., & Minim, L. A. (2020). Short convolutional neural networks applied to the recognition of the browning stages of bread crust. *Journal of Food Engineering*, 277, 109916. doi:10.1016/j.jfoodeng.2020.109916
- Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, (str. 248–255). doi:10.1109/CVPR.2009.5206848
- Donahue, J., Hendricks, L. A., Rohrbach, M., Venugopalan, S., Guadarrama, S., Saenko, K., & Darrell, T. (2017). Long-Term Recurrent Convolutional Networks for Visual Recognition and Description. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39, 677–691. doi:10.1109/TPAMI.2016.2599174
- Gao, L., Guo, Z., Zhang, H., Xu, X., & Shen, H. T. (2017). Video Captioning With Attention-Based LSTM and Semantic Consistency. *IEEE Transactions on Multimedia*, 19, 2045–2055. doi:10.1109/TMM.2017.2729019
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (str. 770–778). doi:10.1109/CVPR.2016.90
- Hochreiter, S., & Schmidhuber, J. (December 1997). Long Short-term Memory. *Neural computation*, 9, 1735–80. doi:10.1162/neco.1997.9.8.1735
- Krizhevsky, A., Sutskever, I., & Hinton, G. (2012, January). ImageNet Classification with Deep Convolutional Neural Networks. *Neural Information Processing Systems*, 25. doi:10.1145/3065386
- Meenu, M., Kurade, C., Neelapu, B. C., Kalra, S., Ramaswamy, H. S., & Yu, Y. (2021). A concise review on food quality assessment using digital image processing. *Trends in Food Science & Technology*, 118, 106–124. doi:10.1016/j.tifs.2021.09.014
- Mogol, B. A., & Gökmen, V. (May 2014). Computer vision-based analysis of foods: A non-destructive colour measurement tool to monitor quality and safety. *J. Sci. Food Agric.*, 94, 1259–1263. doi:10.1002/jsfa.6500

- motion-project. (2020). Motion project. *Motion project*. Pridobljeno iz <https://motion-project.github.io/>
- Ng, J., Hausknecht, M., Vijayanarasimhan, S., Vinyals, O., Monga, R., & Toderici, G. (June 2015). *Beyond short snippets: Deep networks for video classification*. doi:10.1109/CVPR.2015.7299101
- O'Shea, K., & Nash, R. (November 2015). An Introduction to Convolutional Neural Networks. *ArXiv e-prints*.
- Paquet-Durand, O., Solle, D., Schirmer, M., Becker, T., & Hitzmann, B. (2012). Monitoring baking processes of bread rolls by digital image analysis. *Journal of Food Engineering*, *111*, 425–431. doi:10.1016/j.jfoodeng.2012.01.024
- Purlis, E. (2010). Browning development in bakery products - A review. *Journal of Food Engineering*, *99*, 239–249. doi:10.1016/j.jfoodeng.2010.03.008
- Purlis, E., & Salvadori, V. O. (2009). Modelling the browning of bread during baking. *Food Research International*, *42*, 865–870. doi:10.1016/j.foodres.2009.03.007
- Pylar, E. J., & Gorton, L. A. (2008). *Baking science & technology: volume I: fundamentals & ingredients*. Sosland Pub.
- Rudén, C. (2004). Acrylamide and cancer risk—expert risk assessments and the public debate. *Food and Chemical Toxicology*, *42*, 335–349. doi:10.1016/j.fct.2003.10.017
- Shafiee, S., Minaei, S., Moghaddam-Charkari, N., & Barzegar, M. (2014). Honey characterization using computer vision system and artificial neural networks. *Food Chemistry*, *159*, 143–150. doi:10.1016/j.foodchem.2014.02.136
- Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W. K., & Wang-chun, W. O. (June 2015). Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting.
- Tan, M., & Le, Q. (May 2019). *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks*.
- Ureta, M. M., Olivera, D. F., & Salvadori, V. O. (2014). Baking of muffins: Kinetics of crust color development and optimal baking time. *Food and Bioprocess Technology*, *7*, 3208–3216. doi:10.1007/s11947-014-1292-z
- Ureta, M., Olivera, D., & Salvadori, V. (August 2016). Influence of baking conditions on the quality attributes of sponge cake. *Food Science and Technology International*, *23*. doi:10.1177/1082013216666618
- Yadav, S. K., Tiwari, K., Pandey, H. M., & Akbar, S. A. (2022). Skeleton-based human activity recognition using ConvLSTM and guided feature learning. *Soft Computing*, *26*, 877–890. doi:10.1007/s00500-021-06238-7
- Zhu, L., Spachos, P., Pensini, E., & Plataniotis, K. N. (2021). Deep learning and machine vision for food processing: A survey. *Current Research in Food Science*, *4*, 233–249. doi:10.1016/j.crfs.2021.03.009

ZAZNAVANJE OSLONILNIH TOČK AORTNE ZAKLOPKE V CT SLIKAH S POSTOPKOM GLOBOKEGA UČENJA

LUKA ŠKRLJ,¹ MATIJA JELENC,² FRANJO PERNUŠ,¹
TOMAŽ VRTOVEC¹

¹ Univerza v Ljubljani, Fakulteta za elektrotehniko, Ljubljana, Slovenia
luka25.skrlj1@gmail.com, franjo.pernus@fe.uni-lj.si, tomaz.vrtovec@fe.uni-lj.si

² Univerzitetni klinični center Ljubljana, Kirurška klinika, Klinični oddelek za kirurgijo
srca in ožilja, Ljubljana, Slovenia
matija.jelenc@kclj.si

Sinopsis Morfologija aortne zaklopke se v kardiovaskularni medicini uporablja za analizo bolezni srčnih zaklopk, za katero pa je potrebna natančna določitev značilnih anatomske točk, tj. oslonilnih točk. V tej študiji preučujemo uporabo arhitekture globokega učenja, in sicer omrežje prostorskih konfiguracij, za zaznavanje oslonilnih točk na lističih zdrave aortne zaklopke v 120 koronarnih slikah, pridobljenih s tehniko računalniške tomografije. Na podlagi trikratnega prečnega preverjanja smo za šest oslonilnih točk, ki predstavljajo baze in komisure sinusov aortne zaklopke, z omenjeno metodo dobili povprečno napako zaznavanja $1,45 \pm 0,82$ mm. Čeprav je dobljena natančnost primerljiva z obstoječimi metodami, bi lahko dodatne izboljšave v obliki predhodne oz. naknadne obdelave slik ali z uporabo naprednejših metodoloških konceptov izboljšale uspešnost zaznavanja oslonilnih točk.

Ključne besede:

zaznavanje
oslonilnih točk,
globoko učenje,
aortna zaklopka,
morfometrija,
slike CT

DEEP LEARNING-BASED AORTIC VALVE LANDMARK DETECTION IN CT IMAGES

LUKA ŠKRLJ,¹ MATIJA JELENC,² FRANJO PERNUŠ,¹
TOMAŽ VRTOVEC¹

¹ University of Ljubljana, Faculty of Electrical Engineering, Ljubljana, Slovenia
luka25.skrjl1@gmail.com, franjo.pernus@fe.uni-lj.si, tomaz.vrtovec@fe.uni-lj.si

² University Medical Center Ljubljana, Division of Surgery, Department of
Cardiovascular Surgery, Ljubljana, Slovenia
matija.jelenc@kclj.si

Abstract Aortic valve morphology is used in cardiovascular medicine to analyze the valvular heart disease, however, it requires an accurate identification of specific anatomical points, i.e. landmarks. In this study, we applied a deep learning method known as the spatial configuration network for the detection of landmarks on normal aortic cusps in 120 coronary computed tomography images. By performing three-fold cross-validation, we obtained a 1.45 ± 0.82 mm mean detection error for six landmarks representing the nadirs and commissures of the aortic valve sinuses. Although the obtained accuracy is comparable to existing methods, additional improvements in the form of image pre/postprocessing or by relying on more advanced methodological concepts may result in an increase in the performance of landmark detection.

Keywords:

landmark
detection,
deep learning,
aortic valve,
morphometry,
CT images

1 Uvod

Bolezni srčne zaklopke so vse bolj razširjene zaradi boljšega diagnosticiranja ter staranja prebivalstva, zato je njihovo zdravljenje pomembna tema kardiovaskularne medicine (Coffey in dr., 2021). Sodobno kirurško zdravljenje aortnih zaklopk zahteva temeljito morfološko razumevanje njihove tridimenzionalne (3D) anatomije (Tretter in dr., 2021), za morfološko analizo pa se kot dopolnilna tehnika slikovnega zajema uveljavlja predvsem računalniška tomografija (CT) (Tretter in dr., 2021). Izhodišče za morfologijo aortne zaklopke je natančna določitev značilnih anatomskih točk, tj. oslonilnih točk, kot so baze in komisure sinusov aortnih lističev (slika 1(a)). Ker je njihovo ročno določanje na CT slikah težavno, lahko samodejna analiza medicinskih slik zagotovi natančnejšo in zanesljivejšo morfološko analizo. Napredek na področju globokega učenja za samodejno analizo medicinskih slik je vplival tudi na področje kardiovaskularne medicine, na primer, konvolucijske nevronske mreže (CNN) (Yu in dr., 2021) so že bile uporabljene tudi za samodejno zaznavanje oslonilnih točk aortne zaklopke (Noothout in dr., 2020; Aoyama in dr., 2022). V tej študiji bomo obravnavali uporabo CNN arhitekture (Payer in dr., 2019) za zaznavanje oslonilnih točk aortne zaklopke v koronarnih CT slikah.

2 Metodologija

Za zaznavanje oslonilnih točk aortne zaklopke smo uporabili CNN arhitekturo, znano pod imenom omrežje prostorskih konfiguracij (Payer in dr., 2019), ki temelji na predstavitvi vsake oslonilne točke s porazdelitvijo verjetnosti, tj. toplotno sliko, in oblikovanju prostorskih konfiguracij med oslonilnimi točkami (slika 1(b)).

2.1 Toplotne slike oslonilnih točk

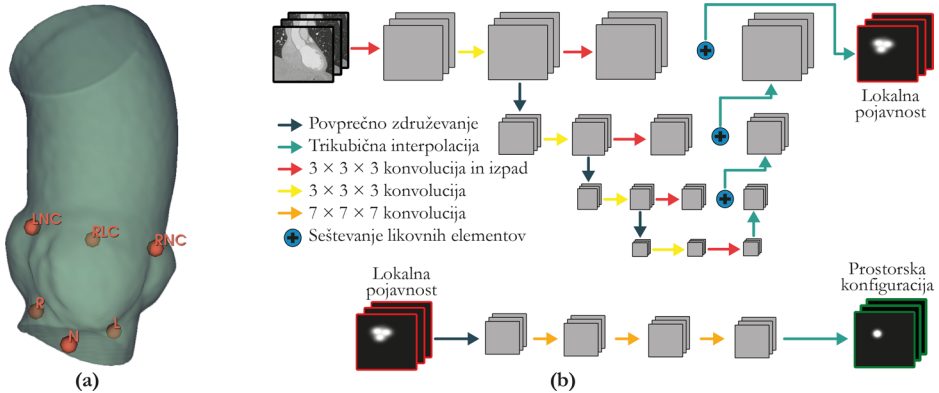
Najprej ustvarimo 3D toplotno sliko g_i za vsako i -to oslonilno točko na vsaki sliki, ki sledi Gaussovi porazdelitvi in predstavlja verjetnost, da se ta točka nahaja na trenutnem položaju (Payer in dr., 2019):

$$g_i(\mathbf{x}, \sigma_i) = \frac{\gamma}{(\sigma_i \sqrt{2\pi})^3} e^{-\frac{\|\mathbf{x} - \hat{\mathbf{x}}_i\|^2}{2\sigma_i^2}}, \quad (1)$$

kjer sta $\mathbf{x} = (x, y, z)$ in $\mathbf{x}_i = (x_i, y_i, z_i)$ poljubne in referenčne 3D koordinate i -te oslonilne točke, $\|\cdot\|$ je evklidska razdalja, σ_i je standardni odklon Gaussove funkcije, γ pa parameter povečave. Za izdelavo toplotnih slik nato naučimo omrežje U-Net (Ronneberger in dr., 2015), pri čemer kot parameter uporabimo σ_i . Omrežje se nauči sočasno regresirati vse toplotne slike z minimizacijo izgube L med napovedano in referenčno toplotno sliko h_i oz. g_i :

$$L(\mathbf{w}, \boldsymbol{\sigma}) = \sum_{i=1}^N \|h_i(\mathbf{x}, \mathbf{w}) - g_i(\mathbf{x}, \sigma_i)\|^2 + \alpha \|\boldsymbol{\sigma}\|^2 + \lambda \|\mathbf{w}\|^2, \quad (2)$$

kjer je N število oslonilnih točk, $\boldsymbol{\sigma} = \{\sigma_i\}$ je vektor Gaussovih standardnih odklonov za vse oslonilne točke (1), \mathbf{w} so uteži omrežja, α oz. λ pa določata pripadajoče prispevke. Napovedane koordinate $\mathbf{x}'_i = (x'_i, y'_i, z'_i)$ za i -to oslonilno točko pridobimo iz lokacije največje vrednosti v h_i .



Slika 1: (a) Oslonilne točke aortne zaklopke: R, L ter N – baza desnega, levega ter nekoronarnega sinusa; RLC, RNC ter LNC – komisura med desnim in levim, desnim in nekoronarnim ter levim in nekoronarnim sinusom. (b) Omrežje prostorskih konfiguracij (Payer in dr., 2019) sestoji iz komponente lokalne pojavnosti in komponente prostorske konfiguracije, z njuno združitvijo pa pridobimo končne toplotne slike.

Vir: lasten.

2.2 Prostorska konfiguracija

Omrežje prostorskih konfiguracij (Payer in dr., 2019) sestoji iz dveh povezanih komponent. Prva komponenta h_i^{LA} modelira lokalno pojavnost oslonilnih točk in zagotavlja lokalno natančne, vendar potencialno dvoumne napovedi, ki so podobne

Gaussovi funkciji v okolici oslonilnih točk. To doseže z večnivojsko strukturo zaporednih konvolucijskih slojev, kjer izhod na vsakem nivoju predstavlja ostanek za naslednji nižji nivo, tako da se vmesne toplotne slike postopoma izboljšujejo ob povečevanju ločljivosti, dokler se ne obnovi osnovna ločljivost. Druga komponenta h_i^{SC} modelira prostorsko konfiguracijo oslonilnih točk in razlikuje med lokalno natančnimi, vendar potencialno dvoumnimi vhodnimi toplotnimi slikami. Prostorske konfiguracije oslonilnih točk so modelirane z vrsto konvolucijskih slojev, ki implicitno vključujejo geometrijski model vseh oslonilnih točk in zagotavljajo odpornost proti napačnemu zaznavanju z omejevanjem odzivov samo na verjetne konfiguracije. Z množenjem obeh komponent po slikovnih elementih pridobimo končno toplotno sliko $h_i = h_i^{LA} \odot h_i^{SC}$ za vsako i -to oslonilno točko.

3 Eksperimenti in rezultati

3.1 Zbirka slik

Uporabili smo 120 koronarnih kontrastnih CT slik v končni diastolični fazi oseb z zdravimi aortnimi zaklopkami. Slike so bile pridobljene s CT napravo Somatom Force (Siemens Healthineers, Erlangen, Nemčija) in imajo velikost $512 \times 512 \times 229\text{--}574$ slikovnih elementov, z velikostjo slikovnega elementa $0,4 \times 0,4 \times 0,6 \text{ mm}^3$. Z uporabo programske opreme Mimics (Materialise NV, Leuven, Belgija) je izkušeni kardiovaskularni kirurg na vsaki sliki ročno označil $N = 6$ anatomskih oslonilnih točk aortne zaklopke (slika 1(a)), tj. baze desnega, levega ter nekoronarnega sinusa (R, L ter N) in komisure med desnim in levim, desnim in nekoronarnim ter levim in nekoronarnim sinusom (RLC, RNC ter LNC).

3.2 Eksperimenti

Uporabili smo javno dostopno kodo⁶ omrežja prostorskih konfiguracij (Payer in dr., 2019). Vrednotenje smo izvedli preko trikratnega prečnega preverjanja, torej smo v vsakem od treh eksperimentov za učenje uporabili 80 slik (66,6 %), za testiranje pa 40 slik (33,3 %). Slike so bile spremenjene na velikost $96 \times 96 \times 128$ slikovnih elementov z velikostjo slikovnega elementa $1 \times 1 \times 1 \text{ mm}^3$, intenzitete pa so bile normalizirane na območje $[-1, 1]$. Bogatenje podatkov smo izvedli z naključnim

⁶ <https://github.com/christianpayer/MedicalDataAugmentationTool-HeatmapRegression>

množenjem oz. premikom intenzitet za $[0,75, 1,25]$ oz. $[-0,25, 0,25]$ ter z naključno translacijo, rotacijo oz. povečavo za $[-25, 25]$ slikovnih elementov, $[-15^\circ, 15^\circ]$ oz. $[0,5, 1,5]$. Uporabljeno metodo globokega učenja smo nato učili preko 80.000 iteracij, napako zaznavanja i -te oslonilne točke pa vrednotili z evklidsko razdaljo $d_i = \|\mathbf{x}'_i - \hat{\mathbf{x}}_i\|$ med napovedanimi in referenčnimi 3D koordinatami \mathbf{x}'_i oz. $\hat{\mathbf{x}}_i$.

Tabela 1: Napaka zaznavanja oslonilnih točk aortne zaklopke (povprečje \pm stand. odklon).

Oslonilna točka	Oznaka	Napaka (mm)
Baza levega sinusa	R	1.39 ± 0.61
Baza desnega sinusa	L	1.32 ± 0.62
Baza nekoronarnega sinusa	N	1.27 ± 0.83
Komisura med levim in desnim sinusom	RLC	1.50 ± 0.62
Komisura med desnim in nekoronarnim sinusom	RNC	1.82 ± 1.11
Komisura med levim in nekoronarnim sinusom	LNC	1.31 ± 0.87
Skupaj		1.45 ± 0.82

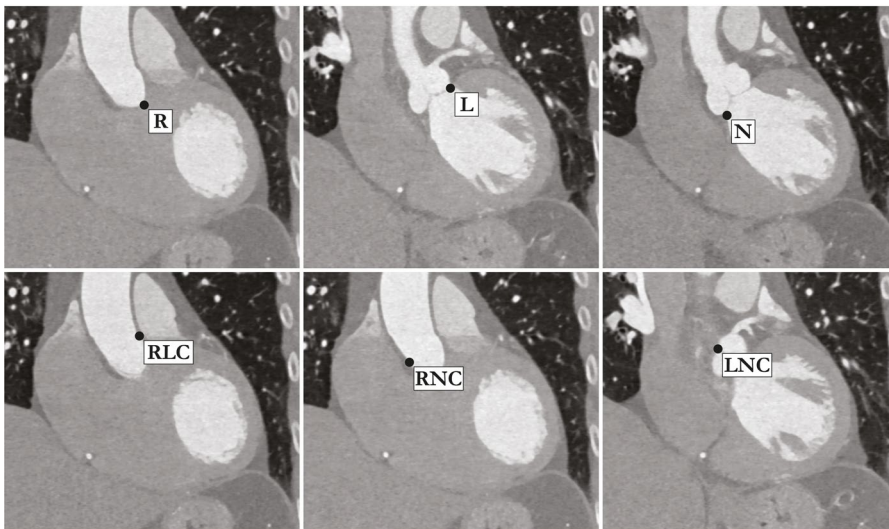
3.3 Rezultati

Z uporabo opisanega omrežja prostorskih konfiguracij (Payer in dr., 2019) za zaznavanje $N = 6$ oslonilnih točk aortne zaklopke smo za 120 koronarnih CT slik ob trikratnem prečnem preverjanju dobili povprečno napako zaznavanja $1,45 \pm 0,82$ mm. Za posamezne oslonilne točke je napaka zaznavanja podana v tabeli 1, medtem ko je primer izbrane koronarne CT slike prikazan na sliki 2.

4 Diskusija

V tej študiji smo uporabili obstoječo arhitekturo za zaznavanje točk aortne zaklopke na koronarnih CT slikah. S hkratno optimizacijo komponent lokalne pojavnosti in prostorske konfiguracije smo problem zaznavanja oslonilnih točk razdelili na dva preprostejša podproblema (Payer in dr., 2019) ter dosegli relativno majhno napako zaznavanja baz in komisur sinusov srčne zaklopke, ki je primerljiva z obstoječimi pristopi na podlagi globokega učenja. Noothout in dr. (2020) so uporabili CNN-je za nadgradnjo približnega z bolj natančnim zaznavanjem, s čimer so dosegli mediano napake zaznavanja baz in komisur $1,87$ mm za 672 koronarnih CT slik. Aoyama in dr. (2022) so isto omrežje prostorskih konfiguracij uporabili v kaskadni obliki, pri čemer so najprej izvedli približno zaznavanje in ga uporabili za segmentacijo aortne zaklopke, nato pa natančneje zaznali oslonilne točke na 138 koronarnih CT slikah

normalnih oseb in bolnikov s kalcifikacijo z napako zaznavanja baz in komisur $1,57 \pm 1,34$ mm. V izvorni študiji o omrežju prostorskih konfiguracij so Payer in dr. (2019) poročali o napaki $0,66 \pm 0,74$ mm pri zaznavanju 37 oslonilnih točk na 895 dvodimenzionalnih rentgenskih slikah dlani, $0,84 \pm 0,62$ mm pri zaznavanju 28 točk na 60 3D magnetno-resonančnih (MR) slikah dlani ter $6,2 \pm 9,9$ mm oz. $2,9 \pm 4,4$ mm pri zaznavanju 26 oslonilnih točk na 224 oz. 60 3D CT slikah hrbtenice. V naši študiji smo dobili povprečno napako zaznavanja $1,45 \pm 0,82$ mm za šest oslonilnih točk v 120 3D koronarnih CT slikah. V primerjavi s študijo Aoyama in dr. (2022) so naši rezultati nekoliko boljši, vendar v naši zbirki slik ni bilo bolezenskih primerov (npr. kalcifikacij). Po drugi strani pa rezultati, o katerih poročajo Payer in dr. (2019), zlasti tisti za 3D MR slike dlani, kažejo na to, da so pri uporabi predlaganega omrežja prostorskih konfiguracij še vedno mogoče izboljšave.



Slika 2: Primer zaznavanja oslonilnih točk v čelnih prerezih CT slike: R, L ter N – baza desnega, levega ter nekoronarnega sinusa; RLC, RNC ter LNC – komisura med desnim in levim, desnim in nekoronarnim ter levim in nekoronarnim sinusom.

Vir: lasten.

Morebitne izboljšave bi se lahko izvedle z nastavitvijo parametrov ali predobdelavo oz. poobdelavo slik, ki v tej študiji ni bila uporabljena. Naše prihodnje delo bo zato osredotočeno na segmentacijo aorte (Chen in dr., 2020), vključitev slik bolezenskih

primerov ter uporabo naprednejših mehanizmov, ki lahko pomagajo pri zaznavanju oslonilnih točk, kot je na primer spodbujevalno učenje (Bekkouch in dr., 2022).

Opombe

To študijo je podprla Javna agencija za raziskovalno dejavnost Republike Slovenije (ARRS) v okviru projekta P2-0232.

Literatura

- Aoyama, G., Zhao, L., Zhao, S., in dr. (2022). Automatic aortic valve cusps segmentation from CT images based on the cascading multiple deep neural networks. *Journal of Imaging*, 8, 11. doi:10.3390/jimaging8010011
- Bekkouch, I.E.I., Maksudov, B., Kiselev, S., Mustafae, T., Vrtovec, T., Ibragimov, B. (2022). Multi-landmark environment analysis with reinforcement learning for pelvic abnormality detection and quantification. *Medical Image Analysis*, 78, 102417. doi:10.1016/j.media.2022.102417
- Chen, C., Qin, C., Qiu, H., in dr. (2020). Deep learning for cardiac image segmentation: a review. *Frontiers in Cardiovascular Medicine*, 7, 25. doi:10.3389/fcvm.2020.00025
- Coffey, S., Roberts-Thomson, R., Brown, A., in dr. (2021). Global epidemiology of valvular heart disease. *Nature Reviews Cardiology*, 18, 853–864. doi:10.1038/s41569-021-00570-z
- Noothout, J.M.H., de Vos, B.D., Wolterink, J.M., in dr. (2020). Deep learning-based regression and classification for automatic landmark localization in medical images. *IEEE Transactions on Medical Imaging*, 39, 4011–4022. doi:10.1109/TMI.2020.3009002
- Payer, C., Štern, D., Bischof, H., Urschler, M. (2019). Integrating spatial configuration into heatmap regression based CNNs for landmark localization. *Medical Image Analysis*, 54, 207–219. doi:10.1016/j.media.2019.03.007
- Ronneberger, O., Fischer, P., Brox, T. (2015). U-Net: Convolutional neural networks for biomedical image segmentation. V: *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015*, LNCS, 9351, 234–241 (Springer). doi:10.1007/978-3-319-24574-4_28
- Tretter, J.T., Izawa, Y., Spicer, D.E., in dr. (2021). Understanding the aortic root using computed tomographic assessment: a potential pathway to improved customized surgical repair. *Circulation: Cardiovascular Imaging*, 14, e013134. doi:10.1161/CIRCIMAGING.121.013134
- Yu, H., Yang, L.T., Zhang, Q., Armstrong, D., Deen, M.J. (2021). Convolutional neural networks for medical image analysis: state-of-the art, comparisons, improvement and perspectives. *Neurocomputing*, 444, 92–110. doi:10.1016/j.neucom.2020.04.157

ANALIZA GEOMETRIJE LIJAKASTIH PASTI V PESKU

TINA KLENOVŠEK,¹ DUŠAN DEVETAK,² JAN PODLESNIK¹

¹ Univerza v Mariboru, Fakulteta za naravoslovje in matematiko, Maribor, Slovenija
tina.klenovsek@um.si, jan.podlesnik@um.si

² Maribor, Slovenija
dusan.devetak@guest.arnes.si

Sinopsis Volkci lijakarji so žuželke, katerih ličinke v pesku gradijo pasti v obliki lijaka za lov plena. Naklon lijaka in sipanje peska proti dnu lijaka plenu otežujeta pobeg, zato je oblika lijaka ključnega pomena za uspeh pri lovu. Različne vrste volkcev lijakarjev lahko gradijo različno oblikovane pasti. V raziskavi Devetak in sodelavci (2020) smo proučevali povezavo med geometrijo pasti in lastnostmi peščenega substrata. Pasti smo 3D skenirali z laserskim čitalnikom in izdelali 3D modele pasti, ki smo jih uporabili za opis oblike z meritvami kotov in za analizo variabilnosti pasti z metodami geometrijske morfometrije. Ugotovili smo, da substrat z manjšo granulacijo ali večjim deležem finega peska omogoča gradnjo kompleksnejših pasti s strmějšími stenami, saj ima višji maksimalni kot stabilnosti. Oblika pasti je torej odvisna od strukture substrata in v heterogenem habitatu omogoča sobivanje različnih vrst volkcev z različnimi preferencami glede substrata.

Ključne besede:

3D laserski
čitalnik,
geometrijska
morfometrija,
analiza oblike,
ličinke volkcev,
granulacija peska

ANALYSIS OF GEOMETRY OF PIT-FALL TRAPS IN SAND

TINA KLENOVŠEK,¹ DUŠAN DEVETAK,² JAN PODLESNIK¹

¹ University of Maribor, Faculty of Natural Sciences and Mathematics, Maribor, Slovenia,

tina.klenovsek@um.si, jan.podlesnik@um.si

² Maribor, Slovenija
dusan.devetak@guest.arnes.si

Abstract Pit-building antlions are insects whose larvae catch prey by constructing pitfall traps in the sand. The slope of the trap and sand avalanches impede the escape of the prey, therefore, the shape of trap is crucial for hunting success. Different species of pit-building antlions can build differently shaped traps. In the study by Devetak et al. (2020), we analysed the connection between the trap geometry and the properties of the sand substrate. We scanned the traps in 3D with a laser scanner and created 3D models of the traps, which we used to describe the shape with angle measurements and to analyse the trap variability using the geometric morphometry methods. We found that a substrate with smaller granulation or a larger proportion of fine sand enables the construction of more complex traps with steeper walls, as it has a higher maximum angle of stability. The shape of the trap therefore depends on the structure of the substrate and in a heterogeneous habitat enables the coexistence of different antlion species with different substrate preferences.

Keywords:

3D laser scanner,
geometric
morphometrics,
shape analysis,
antlion larvae,
sand granulation

1 Uvod

Volkci (Myrmeleontidae) so žuželke iz skupine pravih mrežekrilcev (Neuroptera). Imajo popolno preobrazbo in plenilsko ličinko. Ličinke volkcev na plen prežijo iz zasede. Na plen čakajo zakopane pod površjem substrata ali na dnu peščenega lijaka, ki služi kot past. Plen so običajno mravlje in drugi talni nevretenčarji. Uporaba lijakastih pasti ličinkam omogoča večjo uspešnost pri lovu in lov večjega plena. Ličinke za gradnjo lijakov najraje izbirajo suh, droben pesek. Naklon lijaka in sipanje peska proti dnu lijaka plenu otežujeta pobeg. Ko plen pade v lijak, ga zagrabijo s čeljustmi, vanj vbrizgajo strupe in prebavne encime ter utekočinjeno vsebino plena posesajo.

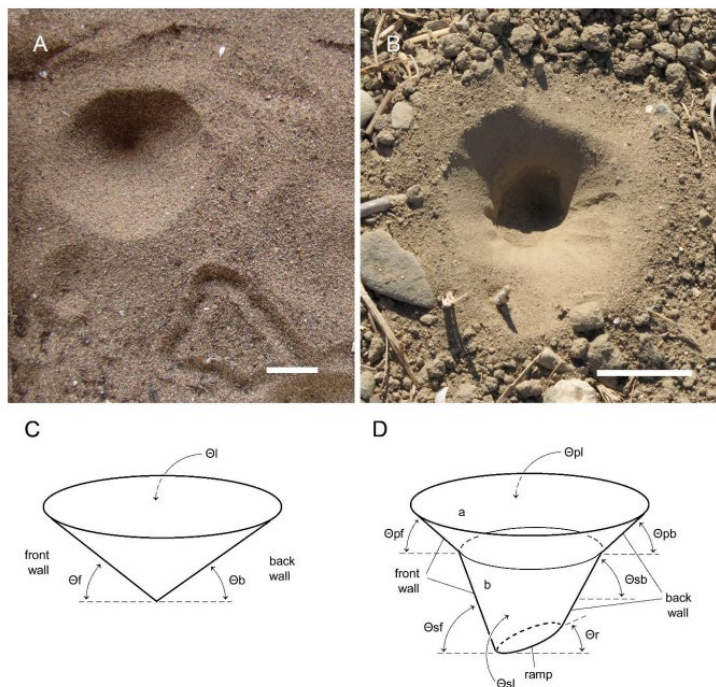
Ličinka volkca koplje lijak v pesek s spiralnim gibanjem nazaj in metanjem peska iz nastajajočega lijaka. Gradnja lijaka predstavlja velik časovni in energetski vložek, zato je učinkovita past ključnega pomena za preživetje ličinke. Ko plen zaide v past, strmina lijaka otežuje pobeg, metanje peska ličinke pa sproži peščeni plaz, ki plen odnese proti dnu (Humeau in sod., 2019). Na uspešnost lova vpliva več dejavnikov: velikost lijaka, naklon sten lijaka, velikost peščenih delcev in fizikalne lastnosti zrnatega medija.

Različne vrste volkcev lahko gradijo različno oblikovane lijakaste pasti, t.j. od pasti v obliki enostavnega narobe obrnjenega stožca do pasti v obliki dveh odsekanih narobe obrnjenih stožcev vstavljenih eden v drugega (Devetak in sod., 2020). Razlike v obliki pasti različnih vrst volkcev so nas vzpodbudile k raziskavi (Devetak in sod., 2020) povezave med geometrijo pasti in značilnostmi substrata, ki ga vrste izberejo za gradnjo. Geometrijo pasti dveh različnih vrst volkcev smo povezali z lastnostmi substrata (velikostjo peščenih delcev, kotom maksimalne stabilnosti in kotom obmirovanja). V tem prispevku predstavljamo tisti del omenjene raziskave, ki se nanaša na analizo 3D modelov peščenih pasti volkcev z metodami geometrijske morfometrije.

2 Ljakaste pasti

Za raziskavo smo uporabili ličinke dveh volkcev *Myrmeleon hyalinus* in *Cueta lineosa*. Prva vrsta gradi enostavnejše pasti v obliki narobe obrnjenega stožca, druga vrsta gradi kompleksnejše pasti v obliki dveh odsekanih narobe obrnjenih stožcev vstavljenih eden v drugega (slika 1A, B). Ličinke smo v Laboratoriju za fiziologijo živali in etologijo na FNM UM gojili v peščenem substratu. Uporabili smo substrat

iz naravnega habitata ličink kot tudi umeten substrat. Izmerili smo lastnosti substratov (velikost delcev, kot obmirovanja (α_r) in maksimalni kot stabilnosti (α_m)), ki so ključne za stabilnost sten lijakov.

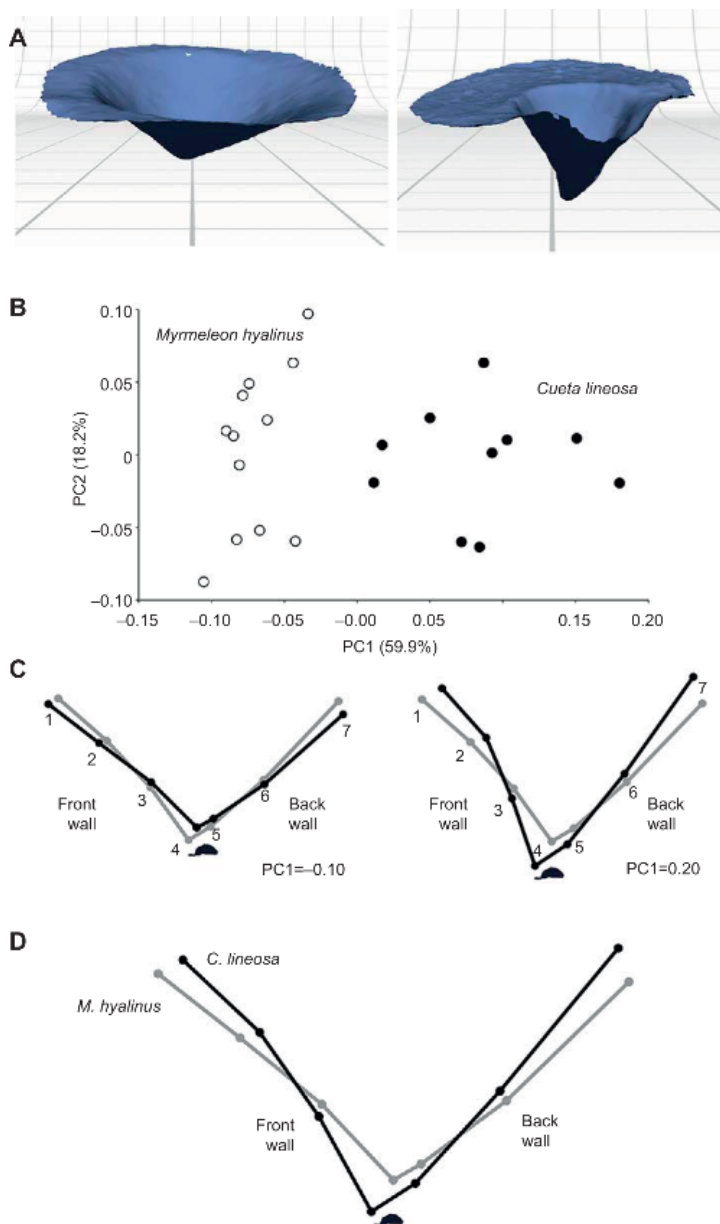


Slika 1: Pasti dveh vrst volkcev (A,C –*Myrmeleon hyalinus*; B, D – *Cueta lineosa*). Za oznake na slikah C in D glej vir: Devetak in sod. 2020

Vir: Devetak in sod. 2020.

2.1 3D skeniranje in modeli lijakastih pasti

Ličinke so v substratu zgradile lijakaste pasti, ki smo jih nato v Laboratoriju za morfometrijo FNM UM (morpholab.fnm.um.si) skenirali s prenosnim 3D laserskim optičnim čitalcem NextEngine 3D Scanner Ultra HD. Optični čitalec smo postavili na podstavek nad posodo s peskom, v kateri je bila lijakasta past. S programom NextEngine ProScan ScanStudio smo izdelali 3D model vsake pasti. Modele pasti smo uporabili za meritve kotov pasti v programu NIS-Elements D 4.20 (slika 1C, D) kot tudi za analizo velikosti in oblike pasti z metodami geometrijske morfometrije (slika 2).



Slika 2: Primerjava oblike pasti dveh vrst volkcev, *Myrmeleon hyalinus* in *Cueta lineosa*. A – 3D model pasti: *M. hyalinus* (levo) in *C. lineosa* (desno), B – razsevni diagram prve (PC1) in druge (PC2) glavne komponente, C – prikaz skrajnih oblik pasti (črni žični graf) v primerjavi s povprečjem (sivi žični graf) na PC1,

D – primerjava povprečne oblike pasti vrste *M. hyalinus* (sivi ž. g.) in *C. lineosa* (črni ž. g.)

Vir: Devetak in sod. 2020.

2.1.1 Analiza pasti z geometrijsko morfometrijo

Geometrijska morfometrija je skupina pristopov za multivariatno statistično analizo podatkov v obliki kartezijskih koordinat pridobljenih z določanjem oslonilnih točk, s katerimi opišemo obliko objektov (Adams in sod., 2013). Metoda omogoča ločeno analizo velikosti in oblike in ves čas analize v celoti ohranjajo vse informacije o relativnih prostorskih razmerjih med oslonilnimi točkami ter s tem omogoča vizualizacijo sprememb in skupinskih ali individualnih razlik v obliki objektov (Klenovšek, 2014).

Za analizo velikosti in oblike pasti z metodami geometrijske morfometrije smo uporabili 2D prereze 3D modelov pasti. Prerezi so bili pravokotni na površino peska in v sredinski ravnini telesa ličink volkcev zakopanih na dnu lijaka (slika 2C, D). Na vsakem 2D prerezu smo določili 7 oslonilnih točk (slika 2C) s programom TpsDig2 (Rohlf, 2015), s katerimi smo opisali obliko pasti. Nato smo točke vseh prerezov poravnali s Procrustovo analizo, ki je konfiguracije točk premaknila v skupno izhodišče, zavrtela in nato skalirala, tako da je bila vsota kvadratov razdalj med točkami na koncu minimalna (Rohlf in Slice, 1990). Tako smo izločili tudi vse razlike med pastmi nastale v procesom digitalizacije in dobili ločene spremenljivke za velikost in obliko pasti. Za zmanjšanje števila oblikovnih spremenljivk in prikaz morfološke variabilnosti pasti smo uporabili analizo glavnih komponent. Za prikaz spreminjanja oblike pasti vzdolž prve glavne komponente in primerjavo povprečnih oblik pasti dveh vrst volkcev smo uporabili žične grafe, ki temeljijo na interpolacijski funkciji TPS (angl. Thin Plate Spline) (Bookstein, 1991). Uporabili smo program MorphoJ (Klingenberg, 2011). Za statistično analizo smo uporabili t-test in analizo variance.

Rezultati analize so pokazali statistično značilne razlike tako v velikosti kot obliki pasti med obema vrstama. Ličinke vrste *M. hyalinus* so gradile večje, a enostavnejše pasti z bolj položnimi stenami. Pasti ličnik vrste *C. lineosa* so bile na splošno manjše, a kompleksnejše, iz dveh obrnjenih stožcev vstavljenih eden v drugega, s širšim zgornjim in ožjim spodnjim delom. Nakloni sten pasti *C. lineosa* so bili različni, tako glede na globino pasti kot lego ličinke (sprednja, zadnja stena, slika 2D). Očitne razlike med pastmi obeh vrst prikazuje tudi graf prvih dveh komponent, kjer vzdolž prve glavne komponente (PC1), ki je razložila skoraj 60% variabilnosti, med posameznimi pastmi ni bilo nobenega prekrivanja.

3 Zaključki

Ličinke *M. hyalinus* so habitatni generalisti, medtem ko so ličinke vrste *C. lineosa* habitatni specialisti (Barkae in sod., 2012). V nadaljevanju raziskave (Devetak in sod., 2020) ličinke *C. lineosa* tako niso gradile pasti v substratu *M. hyalinus* kot tudi ne v umetnem substratu, ki je bil po fizikalnih lastnostih preveč različen od njenega naravnega. Ličinke *M. hyalinus* so bile bolj fleksibilne, saj so gradile pasti v različnih substratih. V substratu, ki je bil podoben naravnemu substratu *C. lineosa*, t.j. finejši pesek z granulacijo $<110 \mu\text{m}$, so tudi ličinke *M. hyalinus* gradile podobne pasti kot *C. lineosa*. Substrat z manjšo granulacijo ali večjim deležem finega peska omogoča gradnjo pasti s strmejšimi stenami, saj ima višji maksimalni kot stabilnosti. Oblika pasti je torej odvisna od strukture substrata. Preferenca in izbira različnega substrata pa vrstama *M. hyalinus* in *C. lineosa* omogoča sobivanje z zmanjšanjem tekmovanja za prostor in plen.

Literatura

- Adams, D. C., Rohlf, F. J., Slice, D. E. (2013) A field comes of age: geometric morphometrics in the 21st century. *Hystrix*, 24(1), 1–8.
- Barkae, E. D., Scharf, I., Abramsky, Z. and Ovidia, O. (2012). Jack of all trades, master of all: a positive association between habitat niche breadth and foraging performance in pit-building antlion larvae. *PLoS ONE* 7, e33506.
- Bookstein, F. L. (1991). Morphometric tools for landmark data. Cambridge, UK: Cambridge University Press.
- Devetak, D., Podlesnik, J., Scharf, I., Klenovšek, T. (2020). Fine sand particles enable antlions to build pitfall traps with advanced three-dimensional geometry. *J Exp Biol*, 223 (15) <https://doi.org/10.1242/jeb.224626>
- Humeau, A., Rougé, J., Casas, J. (2015). Optimal range of prey size for antlions. *Ecol. Entomol.* 40, 776-781. <https://doi.org/10.1111/een.12254>
- Klenovšek, T. (2014). Priročnik za uporabo geometrijske morfometrije v biologiji. Fakulteta za naravoslovje in matematiko. Maribor.
- Klingenberg, C. P. (2011). MorphoJ: an integrated software package for geometric morphometrics. *Mol. Ecol. Res.* 11, 353-357.
- Rohlf, F. J. (2015). The tps series of software. *Hystrix* 26, 9-12.
- Rohlf, F.J., Slice, D. (1990). Extensions of the Procrustes method for the optimal superimposition of landmarks. *Syst. Zool.* 39, 40-59.

ADAPTING VET EDUCATION TO LABOR MARKET NEEDS WITH FOCUS ON ARTIFICIAL INTELLIGENCE AND COMPUTER VISION

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

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

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

² School center Velenje, Velenje, Slovenia

gregor.hrastnik@scv.si, natasa.meh.peer@scv.si

Abstract There is an obvious lack of focus on Artificial Intelligence (AI) in multiple levels of education. The paper presents and is a part of the ongoing EU project AIM@VET (Artificial Intelligence Modules for Vocational Education and Training) that covers the development of learning modules aimed at adapting Vocational Education and Training to the needs of the labor market with a focus on AI. There are six partners from Spain, Portugal and Slovenia, where in the Slovenian branch, the University of Ljubljana (UL) and School Center Velenje (SCV) serve as contributing partners, focusing on computer vision aspect of AI. The topics include capturing and curating unbiased data, detection and segmentation, and tracking and recognition. The development process mainly involves UL preparing the content and the SCV teachers transferring the content to students, with both students and teachers providing feedback. The importance of VET education in AI and the significance of ensuring unbiased and fair algorithms are especially emphasized throughout the project.

Keywords:

EU project
AIM@VET,
computer vision,
artificial
intelligence,
vocational
education and
training,
education

PRILAGAJANJE POKLICNEGA IZOBRAŽEVANJA IN USPOSABLJANJA POTREBAM TRGA DELA Z OSREDOTOČANJEM NA UMETNO INTELIGENCO IN RAČUNALNIŠKI VID

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

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

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

Sinopsis Na večih nivojih izobraževanja je očitno pomanjkanje poudarka na umetni inteligenci (UI). Članek predstavlja zasnovo EU projekta AIM@VET (moduli umetne inteligence za poklicno izobraževanje in usposabljanje) v teku, ki zajema razvoj učnih modulov z namenom prilagajanja poklicnega izobraževanja potrebam trga dela z osredotočenostjo na UI. Projekt vključuje šest partnerjev iz Španije, Portugalske in Slovenije, pri čemer slovenski del predstavljata Univerza v Ljubljani (UL) in Šolski center Velenje (ŠCV), ki se osredotočata na računalniški vid kot delu UI. Teme vključujejo: zajem in urejanje nepristranskih podatkov, detekcijo in segmentacijo ter sledenje in prepoznavanje. Razvojni proces vključuje predvsem pripravo vsebine s strani UL in prenos vsebine s strani učiteljev ŠCV na dijake, pri čemer tako dijaki kot učitelji podajo povratne informacije. Pomen poklicnega izobraževanja na področju UI in pomen zagotavljanja nepristranskih in pravičnih algoritmov sta še posebej izpostavljena skozi projekt.

Ključne besede:

EU projekt
AIM@VET,
računalniški vid,
umetna inteligenca,
poklicno
usposabljanje,
izobraževanje

1 Introduction

The main goal of the project is to address the high need for AI-qualified people in the EU by developing AI teaching units for vocational students. Furthermore, the EU's Digital Education Plan 2021 – 2027 emphasizes that in order to support competitiveness, we need people with the latest digital skills [1]. To face this lack of AI understanding in the educational systems, different initiatives are arising worldwide to develop specific curriculum and literacy for AI, mainly at pre-university level, to prepare students in this field in a formal way [2]. The curriculum will therefore focus on advanced digital topics such as AI and aim to provide the students with up-to-date knowledge and skills that companies are looking for. To achieve this, the VET teachers will be for the delivery and feedback of the teaching units and the focus will be on specific AI topics that are relevant to the market.

The target groups of the project are VET schools, VET teachers, VET students, and technological companies in Spain, Portugal, and Slovenia. The project aims to provide support and training to VET schools, teachers and students in Europe to increase their digital skills and knowledge in advanced digital topics like Artificial Intelligence (AI). The goals are to adapt the curricula and courses to labor market needs, improve the digital skills of teachers and students, and provide practical training in areas like computer vision, robotics, and ambient intelligence. Additionally, the project also aims to increase the interest of VET schools and teachers towards innovation in education and provide better prepared professionals to technological companies in Spain, Portugal, and Slovenia to help them compete in the global digital market. A study [3] reports that by the end of the decade 70% of the companies will have some level of AI integration.

The specific objective is to create 3 learning modules and 12 teaching units for each of the three branches (work packages) of the project, namely computer vision, robotics, and ambient intelligence, where each work package will be covered by two institutions from each country:

- Slovenian work island: University of Ljubljana (UL) and School center Velenje (SCV). Focused on the computer vision module.
- Spanish work island: University of Coruña and Integrated Professional Training Center Rodolfo Ucha Piñeiro. Focused on the robotics module.

- Portuguese work island: University of Minho and Caldas das Taipas Secondary School. Focused on the ambient intelligence module.

Each work package will consist of approximately 140 teaching hours, teacher guides, student guides, programming exercises, and additional materials. The teaching methodology emphasizes hands-on, project-based learning, using standard software libraries and real-world tasks. Open simulation environments and real devices will be used to support online and blended learning. The teachers at partner VET schools will play a crucial role in the design and implementation of the learning modules.

2 Computer Vision

The goal of this work package is to develop teaching units and resources in the field of computer vision for VET students and will be covered by the Slovenian group – UL and SCV.

Computer vision is a crucial aspect of artificial intelligence and ranks fourth among all world publications according to Google Scholar metrics for the IEEE/CVF Conference on Computer Vision and Pattern Recognition publication. It utilizes images as input and extracts information from them to make decisions based on their content. Understanding the concepts, methods, and applications of computer vision is vital for future generations to effectively use and adapt them in daily life.

Today, and even more so in the future, cameras are used in a variety of settings, such as industry facilities, smart homes, autonomous vehicles, surveillance, and entertainment systems. The specific objectives are related to these scenarios and cover topics such as the importance of unbiased image data curation, object detection and segmentation, and tracking and recognition. The teaching units will encompass classical and deep learning solutions, with a focus on different learning techniques, knowledge distillation for lightweight modeling, and explainable decisions.

The teaching units will be hands-on and interconnected, providing examples of real-world applications. Most of the solutions will be based on deep learning and will address the challenges of supervised learning, including a lack of data and labeling, as well as introduce semi- and unsupervised learning and knowledge distillation for

embedded systems such as IoT and robotics. The teaching units will be based on tools such as Python, PyTorch, OpenCV and related tools, with the possible use of real cameras to capture and manipulate real-time video streams and split into the following learning modules (LMs).

2.1 Learning Module 1: Capturing and Curing Unbiased and Properly Distributed Data

In recent years, the topic of bias and fairness in algorithms has become increasingly important due to the potential for automated decision-making models to make unfair decisions with significant societal, legal, and ethical implications. These learning modules will address this crucial issue by presenting a comprehensive protocol for collecting, organizing, labelling, and maintaining image datasets that ensures a balanced distribution of samples and eliminates any potential biases. This will be achieved through teaching units designed to educate students on the proper techniques and methodologies for capturing and curating data in a manner that promotes fairness unbiased decision-making and is within the rules of the General Data Protection Regulation (GDPR).

A case incorporating this will be data for recognition scenario and demonstration of failure/success of the recognition model based on the severe bias present/absent in the training data [4, 5].

2.2 Learning Module 2: Detection and Segmentation

The first step in most Computer Vision systems is object or subject detection, where bounding boxes are used to identify the location of the desired object or subject within an image [6]. However, this technique is limited in that it only provides a rough estimate of the object's location [7]. To overcome this limitation, a more detailed technique called segmentation is also used, which labels each individual pixel as being a part of the object or the background. This learning module will build upon the data obtained in Learning Module 1 and cover both object detection and segmentation in detail, providing students with a comprehensive understanding of these fundamental steps in computer vision systems.

A case incorporating this will include data annotation, which will then be fed into a CNN-based model to locate objects in images. Students will observe the importance of not only raw model capabilities but also the importance of training data.

2.3 Learning Module 3: Tracking and Recognition

While processing a single image is a crucial starting point, many real-world applications require a series of images to be processed within a video, providing an additional temporal component. This learning module will show students how to utilize this extra information in a tracking scenario [10] and demonstrate how it can be used to track objects and subjects over time. The goal of a computer vision system is often recognition, which can be used for a variety of purposes, including the recognition of people, objects, soft-biometrics modalities, gestures, etc. [8, 9] This learning module will provide students with an in-depth understanding of recognition techniques and how they can be applied to real-world problems. By the end of this module, students will have a comprehensive understanding of tracking and recognition techniques and be well-equipped to implement these techniques in their own projects.

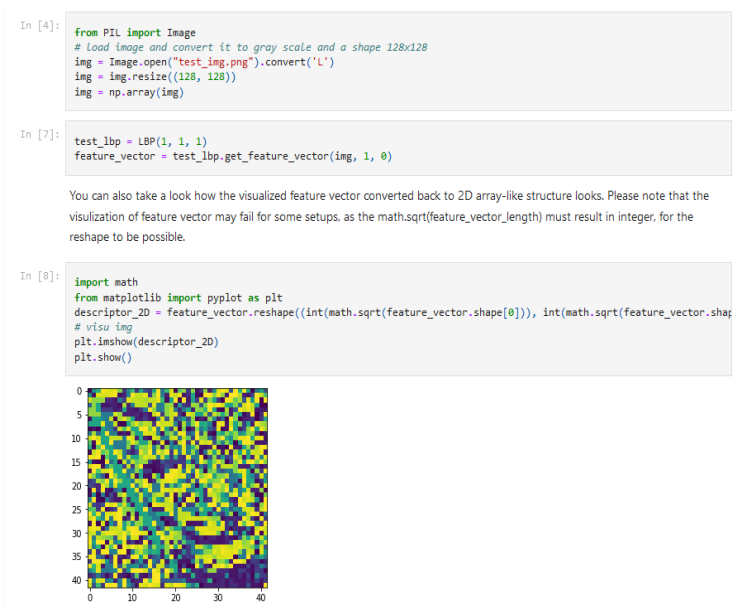


Figure 1: Example of Python Jupyter script within the learning module for feature extraction, which can then be used for image recognition.

Source: own.

A case incorporating this will include developing various feature extraction techniques, as shown in Figure 2, and comparing them to some of the shallower CNN-based architectures. The goal is for students to see that even simple feature extraction techniques work to some extent and can be very useful even with little programming.

3 Expected Results

The University of Ljubljana (UL) and School Center Velenje (SCV) will work together to develop the teaching units (TUs) with a focus on computer vision to contribute to the adaptation of VET education to the labor market needs. The following sequence of tasks will be followed in the development of TUs, as also illustrated in Figure 2:

- UL, in collaboration with SCV, will conduct an initial analysis of the content and the most appropriate means for practical learning.
- UL will implement the TUs for the different learning modules in a sequential manner. On average, each TU will take around 2 months for the UL team to deliver.
- SCV teachers will implement the TUs with their student group as they are received from UL. The implementation time will depend on the VET school's dedication, but a minimum of 2 hours per week is expected.
- SCV teachers and UL team will have on-demand development meetings to clarify any issues and make improvements to the TUs.
- SCV students and teachers will provide feedback on the implemented TUs.
- UL will integrate agreed modifications and deliver the final version of the TUs.

Both partners will work in parallel throughout the project, with a permanent communication channel to ensure a fluid collaboration. The UL team has the technical responsibility, utilizing their expertise in computer vision research and teaching, while SCV has the educational responsibility, bringing their expertise in VET education requirements.

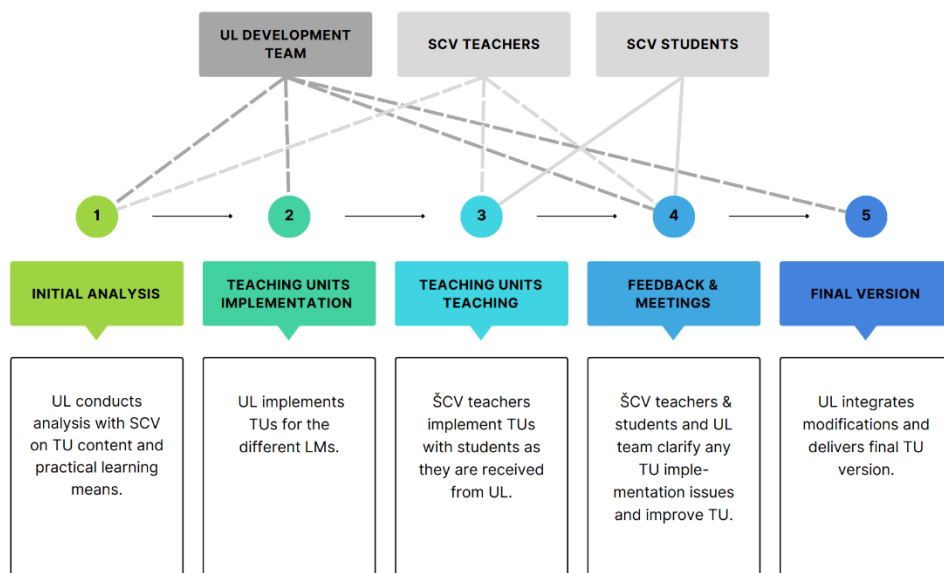


Figure 2: Implementation pipeline.

Source: own.

4 Conclusion

The partnership between, not only the University of Ljubljana and School Center Velenje, but also Spanish and Portuguese partners, presents a unique opportunity to bridge the gap between academia and industry by adapting vocational education and training to meet the needs of the labor market with a focus on AI. We hope that the collaboration will provide students with practical and hands-on experience in capturing, curating, utilizing data, and understanding computer vision algorithms through a series of teaching units. The focus on eliminating biases and promoting fair algorithms is especially important given the potential societal and ethical implications of these systems. The provision of feedback and on-demand development meetings ensures that the teaching units are continuously improved to meet the evolving needs of the field. This innovative approach to education highlights the crucial role that education plays in ensuring that the development and deployment of AI is ethical and responsible.

Acknowledgement

This work was supported in part by the EU Erasmus+ project KA220-VET and ARRS project P2-0214 (B) Computer Vision.

Literature

- [1] EC-European Commission. "Digital Education Action Plan (2021-2027)." (2021). Source: <https://education.ec.europa.eu/focus-topics/digital-education/action-plan>
- [2] Miao, Fengchun, et al. *AI and education: A guidance for policymakers*. UNESCO Publishing, 2021.
- [3] Bughin, Jacques, et al. "Notes from the AI frontier: Modeling the impact of AI on the world economy." *McKinsey Global Institute* 4 (2018).
- [4] Drozdowski, Pawel, et al. "Demographic bias in biometrics: A survey on an emerging challenge." *IEEE Transactions on Technology and Society* 1.2 (2020): 89-103.
- [5] Emeršič, Žiga, et al. "Evaluation and analysis of ear recognition models: performance, complexity and resource requirements." *Neural computing and applications* 32 (2020): 15785-15800.
- [6] Wang, Risheng, et al. "Medical image segmentation using deep learning: A survey." *IET Image Processing* 16.5 (2022): 1243-1267.
- [7] Emeršič, Žiga, et al. "Contextednet: Context-aware ear detection in unconstrained settings." *IEEE Access* 9 (2021): 145175-145190.
- [8] Singh, Maneet, et al. "Recognizing disguised faces in the wild." *IEEE Transactions on Biometrics, Behavior, and Identity Science* 1.2 (2019): 97-108.
- [9] Štepec, Dejan, et al. "Constellation-based deep ear recognition." *Deep biometrics* (2020): 161-190.
- [10] Lukezic, Alan, et al. "Trans2k: Unlocking the Power of Deep Models for Transparent Object Tracking." *arXiv preprint arXiv:2210.03436* (2022).

GENERIRANJE SINTETIČNIH SLIK IZ 3D MODELOV PREDMETOV ZA UČENJE SEGMENTACIJSKIH MODELOV V INDUSTRIJSKEM OKOLJU

MATIC FUČKA, MARKO RUS, JAKOB BOŽIČ,

DANIJEL SKOČAJ

Univerza v Ljubljani, Fakulteta za računalništvo in informatiko, Ljubljana, Slovenija
mf4782@student.uni-lj.si, marko.rus@fri.uni-lj.si, jakob.bozic@fri.uni-lj.si,
daniel.skocaj@fri.uni-lj.si

Sinopsis Ena večjih preprek uporabe globokega učenja v industriji je zahteva po velikem številu označenih slik potrebnih za nadzorovano učenje. Ker je pridobivanje označenih slik pogosto zamuden in drag proces, se veliko metod skuša izogniti temu problemu z bogatenjem učnih slik, ki so na voljo, ali z uporabo sintetičnih slik za predhodno učenje in poznejše doučenje z resničnimi slikami. V tem članku predlagamo metodo za generiranje sintetičnih slik iz danih 3D CAD modelov in novo podatkovno množico sestavljeno iz sintetičnih in resničnih slik ter njihovih segmentacijskih slik. Obravnavamo torej scenarij učenja segmentacijskega modela zgolj iz sintetičnih slik, ki so v industriji tipično na voljo, in zagotavlja najhitrejšo prilagoditev proizvodnega procesa na nove izdelke brez potrebe po zajemanju pravih učnih slik. Na predlagani podatkovni množici ovrednotimo algoritem za segmentacijo slik in primerjamo rezultate dobljene z različnim številom sintetičnih in realnih slik industrijskega predmeta posnetega oz. upodobljenega na različnih ozadjih.

Ključne besede:

strojni vid,
globoko učenje,
segmentacija,
sintetične slike,
CAD model

3D-MODEL-BASED RENDERING OF SYNTHETIC IMAGES FOR TRAINING SEGMENTATION MODELS IN AN INDUSTRIAL ENVIRONMENT

MATIC FUČKA, MARKO RUS, JAKOB BOŽIČ,
DANIJELO SKOČAJ

Univerza v Ljubljani, Faculty of Computer and Information Science, Ljubljana, Slovenia
mf4782@student.uni-lj.si, marko.rus@fri.uni-lj.si, jakob.bozic@fri.uni-lj.si,
danijel.skocaj@fri.uni-lj.si

Abstract One of the major obstacles to the application of deep learning in industry is the requirement for a large number of labeled images required for supervised learning. This is because obtaining labeled images can be time-consuming and costly. To overcome this challenge, some methods utilize image augmentation or synthetic images for pre-training, followed by fine-tuning with real images. This paper introduces a method for generating synthetic images from 3D CAD models, along with a new dataset consisting of both synthetic and real images, and their corresponding segmentation masks. The aim is to train a segmentation model using only synthetic images, which are readily available in industry, allowing for a quicker adaptation of the production process to new products without the need for capturing real training images. We evaluate an image segmentation algorithm on the proposed dataset and compare the results obtained with a different number of synthetic and real images of an industrial object captured or rendered on different backgrounds.

Keywords:
machine vision,
deep learning,
segmentation,
synthetic images,
CAD model

1 Uvod

Področje računalniškega vida je na krilih globokega učenja v zadnjem desetletju doseglo ogromen napredek. Rešitve, temelječe na globokih nevronskih mrežah, vse bolj prodirajo tudi v sisteme strojnega vida, ki so nepogrešljiv del modernih proizvodnih linij v paradigmi Industrija 4.0. Omogočajo namreč veliko bolj raznovrstno in robustno uporabo in rešitev veliko bolj kompleksnih problemov kot do zdaj prevladujoče klasične rešitve strojnega vida.

Ima pa tak, na podatkih in strojnem učenju temelječ pristop, tudi veliko pomanjkljivost: za delovanje namreč zahteva veliko število učnih slik. Še več, trenutno prevladujoč pristop v zrelih aplikacijah strojnega vida temelji na nadzorovanem učenju, ki zahteva, da so učne slike tudi ustrezno označene. Poleg zajemanja slik je torej potrebo tudi veliko človeškega dela za njihovo označevanje. To pa proces uvajanja tovrstnih sistemov v industrijsko proizvodnjo in prilagajanje na nove izdelke podaljša in podraži.

V tem članku obravnavamo eno izmed možnosti za obravnavanje tega problema – umetno generiranje sintetičnih učnih slik z uporabo 3D CAD modelov. V industrijski proizvodnji so 3D modeli izdelkov tipično na voljo, tudi samo okolje je običajno dokaj poznano. To predstavlja zelo ugodne pogoje za generiranje sintetičnih slik, katerih pojavnost ne odstopa bistveno od pojavnosti pravih slik posnetih v pravem okolju. Je pa tako generiranje sintetičnih slik, ob ustrezno razviti programski opremi, hitro in praktično zastonj. V kolikor bi bile zgenerirane učne slike dovolj realistične, bi torej tak pristop pomenil idealen scenarij za gradnjo globokih modelov za reševanje nalog strojnega vida v industrijskem okolju.

V nadaljevanju bomo predstavili cevovod za sintetično generiranje slik ter tako zgenerirano učno množico slik. Osredotočili se bomo na konkreten problem segmentacije predmeta oz. ločevanja pikslov, ki pripadajo ospredju (tj. predmetu), od tistih, ki pripadajo ozadju. To je namreč običajno pogosto prvi problem, ki ga mora sistem strojnega vida rešiti. Analizirali bomo rezultate, dobljene z učenjem na sintetičnih slikah in jih primerjali z rezultati, dobljenimi z uporabo pravih učnih slik. Eksperimentalno bomo torej raziskali uporabnost zgeneriranih sintetičnih učnih slik za namen segmentacije predmetov v realnem industrijskem okolju.

2 Sorodna dela

Uporaba sintetičnih podatkov za namene globokega učenja je vedno bolj priljubljena. To še posebej velja za področje računalniškega vida. V preteklih nekaj letih je bilo ustvarjeno mnogo podatkovnih množic za reševanje različnih problemov kot so strojno štetje (Wang et al., 2019), za namene avtonomne vožnje (Richter et al., 2016; Li et al., 2019; Huang et al., 2018; Yogamani et al., 2019; Zhan et al., 2019) ter razpoznavo raznih detajlov samih slik (McCormac et al., 2017; Mayer et al., 2016). V zadnjem letu ali dveh se je pričelo generirati veliko sintetičnih slik ustvarjenih z raznimi generativnimi modeli kot so npr. generativne nasprotniške mreže (ang. generative adversarial networks) (Dewi et al., 2022) in difuzijski modeli (Pinaya et al., 2022; Akrouf et al., 2023) v nasprotju z obdobjem pred tem, ko se je večino sintetičnih slik ustvarilo s programi za grafično upodabljanje. Večina metod (Li et al., 2019) uporablja sintetične slike kot podaljšek učne množice, ki je sestavljena samo iz resničnih slik, nekaj pa jih zgradi celotno učno množico iz sintetičnih slik (Mayershofer et al., 2021; Mayer et al., 2016) in nato naredi premik domene z uporabo resničnih slik za doučenje (Wang et al., 2019; Richter et al., 2016) nevronske mreže.

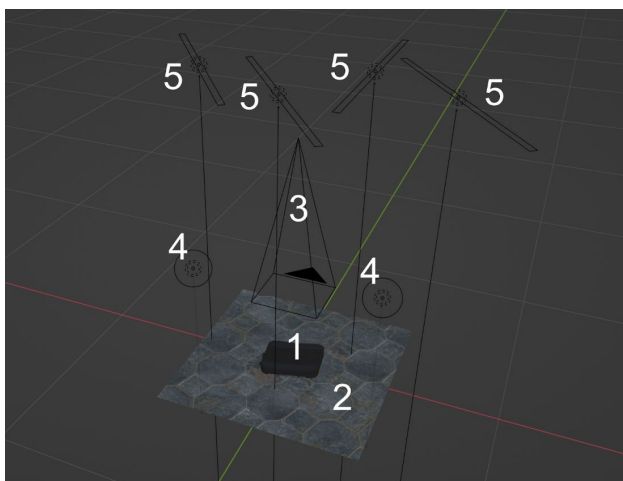
V industrijski domeni (Mayershofer et al., 2021; Eversberg et al., 2021; Abou et al., 2022; Quattrocchi et al., 2022) pogosto pride do uporabe sintetičnih slik, saj je pogosto zajemanje in označevanje velike količine slik drago in zamudno. Velika večina pristopov ustvari sintetične slike na podoben način. Podan imajo 3D CAD model in nato s pomočjo programov za grafično upodabljanje, kot sta Blender in NVidia Omniverse, ustvarijo sintetične slike. Te kasneje uporabijo za učenje ter imajo mogoče označeno še manjšo množico resničnih slik, ki jih uporabijo, da doučijo nevronske mreže. Podoben način obravnavamo tudi v tem delu in analiziramo rezultate na problemu segmentacije slik.

3 Generiranje sintetičnih slik

Sintetično množico za učenje modelov globokega učenja ustvarimo z uporabo namenskih računalniških okolij za sintetično upodabljanje slik. Glavna prednost takšnega pristopa je možnost popolne avtomatizacije celotnega postopka upodabljanja, kar pomeni, da lahko brez človeškega posega upodobimo poljubno število sintetičnih slik skupaj s pripadajočimi segmentacijskimi slikami (ang.

segmentation masks). Kljub sintetični naravi postopka lahko poskrbimo za visoko pestrost končnih slik, menjujemo lahko izgled predmeta in lego predmeta, parametre kamere, podlago, na kateri stoji predmet, ozadje in svetila. Poleg tega pa imajo na tak način pridobljene segmentacijske slike popolno točnost, kar pa ne velja nujno za ročno označene slike, pri katerih se lahko zaradi človeške napake zgodi, da določen slikovni element ne pripada označenemu objektu.

Eno od orodij, ki se uporablja za sintetično ustvarjanje slik, je Blender (Community et al., 2018), ki je brezplačen in odprtokodni program za animacijo, modeliranje in fotorealistično upodabljanje 3D okolja. Skupaj z njim uporabljamo tudi programski pripomoček BlenderProc (Denninger et al., 2019), ki poskrbi za avtomatizacijo celotnega postopka upodabljanja znotraj programskega jezika Python. Z njim lahko programsko nastavimo postavitev ciljnega objekta in kamere, določimo podlago in osvetlitev scene ter nastavimo parametre ustvarjanja slike, ki se nato izvede v programu Blender.



Slika 1: Elementi scene znotraj programskega okolja Blender pri ustvarjanju sintetično ustvarjene podatkovne množice. Scena je sestavljena iz sledečih elementov: 1 - ciljni objekt, 2 - podlaga, 3 - kamera, 4 - okrogli luči in 5 - paličaste luči.

Vir: lasten.

V nadaljevanju bomo podrobno opisali postopek tvorjenja predlagane sintetično ustvarjene podatkovne množice, ki vsebuje 6160 parov sintetično ustvarjenih slik in segmentacij. Od tega jih ima 2160 ozadja različna od teh, ki smo jih imeli pri zajemu

resničnih slik. Na začetku postopka upodabljanja sintetične slike se najprej določi parametre objektov na sceni. Na Sliki 1 lahko vidimo vse elemente, ki sestavljajo sceno, in njihove postavitve: ciljni objekt, podlaga, kamera in različna svetila. Ciljni objekt je postavljen na sredino podlage tako, da je njegova širina v virtualnem okolju približno 0,3 m. Nato se določijo parametri elementov, ki se izberejo naključno iz vnaprej določene domene. Določi se rotacijo ciljnega objekta v katerokoli smer iz ravnine podlage in izbere naključno podlago iz množice javno dostopnih tekstur⁷. Osvetlitev je ustvarjena s pomočjo dveh tipov luči, okroglih in podolgovatih. V polovici primerov se izbere okrogli tip luči, ustvari se dvojje luči na nizki višini 0,8 m, tako da se pojavi visoka verjetnost izrazitejših senc. Polmer luči se nastavi od 0 do 0,1 m, svetilnost pa med 40 do 70 lumnov. V drugi polovici primerov se izbere od ene do štiri podolgovate luči, ki so na višini 1,5 m, vsaka ima svetilnost od 5 do 10 lumnov, postavljene so pa v naključno lokacijo in smer. Ta dva načina sta bila izbrana namensko, da bi povečala pestrost osvetlitve, eden poustvarja pestrost senc, drugi pa odsevov, saj je verjetnost neposrednih odbojev svetlobe iz svetila v senzor zaradi dolžine luči povečana. Kamera je postavljena na višino 1,2 m in ima po širini vidno polje 22,6 stopinj in ločljivost 1824×1536 pikselov.

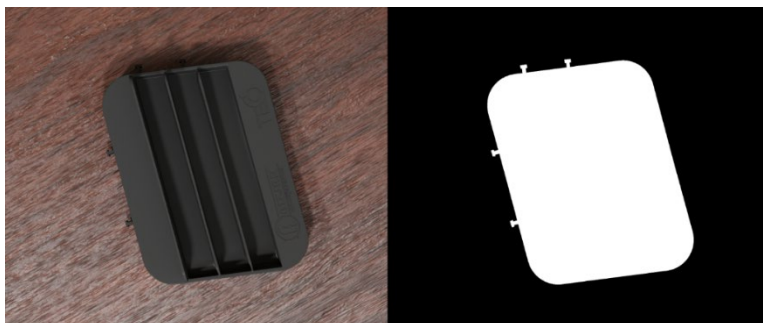


Slika 2: Šest primerov sintetično ustvarjenih slik. Največjo pestrost predstavljata ozadje in rotacija ciljnega objekta, slike se pa razlikujejo tudi v postavitvi svetil in posledično različnih sencah.

Vir: lasten.

⁷ Dostopno na ambientCG.com (licenca CC0).

Slike, ustvarjene z opisanim postopkom, predstavljajo velik nabor možnih osvetlitev in leg predmeta ter podlag, na katerih predmet leži. Raznolikost končnih slik se lahko vidi na Sliki 2, kjer je prikazano nekaj naključnih primerov z opisanim postopkom sintetično ustvarjenih slik. Na Sliki 3 lahko vidimo par sintetično ustvarjene slike in pripadajoče segmentacijske slike, ki se popolnoma prilega predmetu na sliki.



Slika 3: Primer para sintetično ustvarjene slike in segmentacije.

Vir: lasten.

4 Zasnova eksperimenta

V tem razdelku bomo predstavili zasnovo eksperimenta s katerim smo ovrednotili kakovost zgeneriranih učnih slik za namen segmentacije. V ta namen smo posneli množico pravih slik predmeta, s katerim smo preverili delovanje naučenega modela. Zanimala nas je odvisnost uspešnosti zgrajenega modela od števila učnih slik ter tudi od raznovrstnosti ozadja uporabljenega v fazi učenja, pa tudi v fazi vrednotenja.

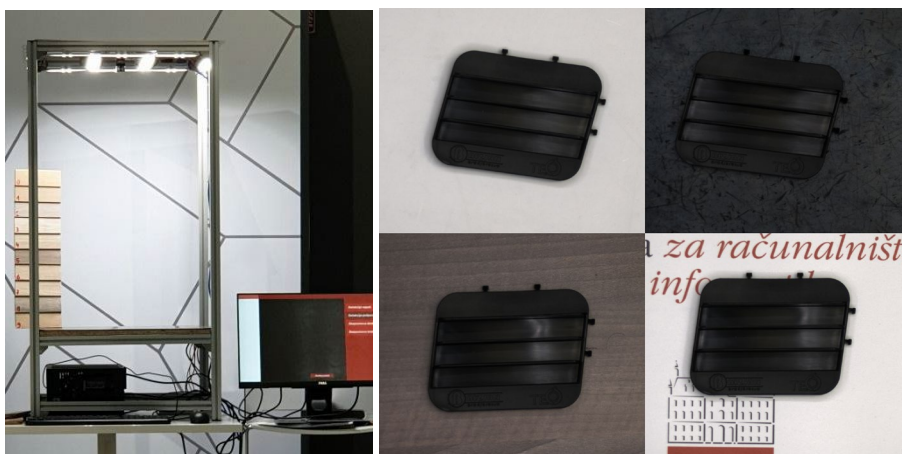
4.1 Zajem resničnih slik

Resnične slike smo zajeli s pomočjo ViCoS Demo celice, ki jo je mogoče videti na Sliki 4. Predmeti na slikah so bili zajeti na štirih različnih ozadjih (belo, črno, leseno in plakat) pri petih različnih osvetlitvah in desetih različnih postavitev predmeta. Belo ozadje je bilo izbrano kot najenostavnejši scenarij na katerem je temen predmet lepo viden. Črno ozadje tako predstavlja zelo težak scenarij saj je kontrast med ozadjem in predmetom (ter njegovimi sencami) zelo majhen. Leseno ozadje predstavlja dokaj realističen scenarij, medtem ko plakat vsebuje bolj nehomogeno

ozadje z nekaj izrazitimi elementi (črke, kvadrat), ki bi utegnili pri segmentaciji povzročati probleme. Nekaj primerov zajetih slik je vidnih na Sliki 4.

Večjo variabilnost scen smo dosegli s spreminjanjem osvetlitve. Na ta način so se na predmetih pojavljali odbleski, nehomogena osvetlitev in tudi sence, ki tudi sicer povzročajo probleme v realnem okolju. Čas ekspozicije smo avtomatsko nastavljali na način, da je bila svetlost prazne scene (slike kalibracijskega lista) podobna ne glede na osvetlitev. Tako smo zagotovili ponovljivost eksperimenta in dovolj velik kontrast na slikah.

Skupaj smo tako posneli 200 pravih slik (10 scen pri 5 osvetlitvah za vsako izmed 4 ozadij). Vse slike smo ročno segmentirali, tako da smo na njih označili piksele, ki pripadajo predmetu.



Slika 4: ViCoS Demo celica in primeri slik predmeta zajetih na vseh štirih ozadijih.

Vir: lasten.

4.2 Postavitev eksperimenta

Imeli smo dva eksperimenta. V prvem eksperimentu smo imeli resnično in sintetično učno množico. Resnična učna množica je bila sestavljena iz 100 resničnih slik, sintetična pa iz 1500 sintetičnih slik. Pri generiranju sintetičnih slik smo uporabljali naključna ozadja. Testno množico predstavlja preostalih 100 resničnih slik. Pri razbitju slik na učno in testno množico smo zagotovili, da so bile vse slike iste scene

(pri različnih osvetlitvah) vedno bodisi v učni bodisi v tesni množici. V tem eksperimentu smo se osredotočili na primerjavo rezultatov dobljenih z učenjem na sintetičnih slikah s tistimi dobljenimi z uporabo pravih slik.

V drugem eksperimentu smo se osredotočili na analizo rezultatov glede na različna ozadja. Zato smo za testno množico vzeli vseh 200 pravih slik (50 za vsako ozadje). Za učno množico smo zgenerirali 800 sintetičnih slik z vnaprej znanimi ozadji, ki so bila podobna resničnim (200 slik za vsako izmed 4 ozadij).

4.3 Mere za vrednotenje uspešnosti

Za vrednotenje uspešnosti segmentacijskega modela smo uporabili meri uspešnosti Povprečna natančnost (ang. Average Precision) in Presek nad unijo (ang. Intersection over union, IoU).

Povprečna natančnost je mera uspešnosti, ki je določena kot ploščina pod krivuljo natančnost-priklic in je neodvisna od postavitve praga ločitve med pikslu ozadja in ospredja.

Presek nad unijo je mera uspešnosti segmentacijskega algoritma, določena kot količnik med velikostjo preseka in unije z algoritmom segmentirane regije in regije, ki je podana kot prava vrednost. Pri binarizaciji segmentacijskega izhoda, torej določanju pikslu ospredja, smo prag postavili na privzeto vrednost 0,5. Običajno se optimalni prag izbere na osnovi nastavitvene množice. Ker pa v scenariju, ki ga obravnavamo v tem članku, pravih učnih in validacijskih slik ni na razpolago, smo uporabili prednastavljeno fiksno vrednost.

4.4 Segmentacijski model

Za vse eksperimente smo uporabili metodo DeepLab++ (Lin et al., 2017). Deeplab++ je globoka konvolucijska nevronska mreža, ki uporablja razširjeno konvolucijo (ang. atrous convolution) na več različnih velikostih slike. To uporablja zato, da bolje zajame globalne značilke na ravni slike. Poleg tega uporablja še vrsto drugih mehanizmov s katerimi na splošnih nalogah semantične segmentacije dosega zelo dobre rezultate in velja za eno izmed najboljših segmentacijskih metod.

Model smo učili skozi 50 epoh z uporabo goriščne izgube (ang. focal loss) (Chen et al., 2017). Uporabili smo optimizator Adam s stopnjo učenja 10^{-4} . Slike smo normalizirali s srednjo vrednostjo in varianco slik dobljenimi na podatkovni množici slik ImageNet (Deng et al., 2009). Med učenjem smo s 50% verjetnostjo sliko malce zameglili in ji spremenili svetlost, odtenek ter nasičenost. Na ta način smo učno množico slik še malce obogatili in povečali njeno raznovrstnost, da bi se izognili prevelikemu prileganju naučenega modela na specifične učne množice.

5 Eksperimentalni rezultati

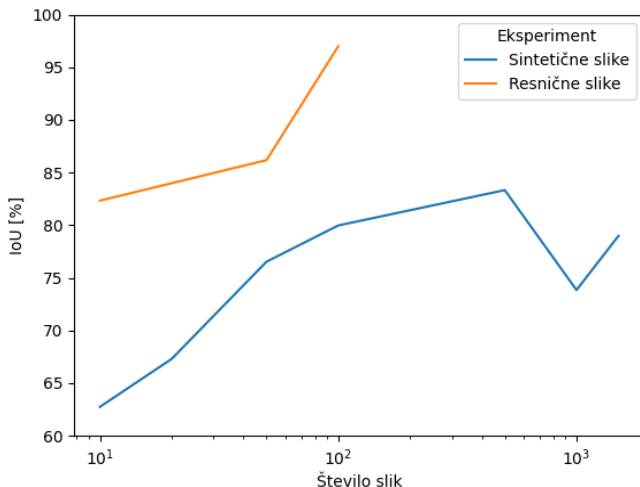
5.1 Prvi eksperiment

V prvem eksperimentu smo opazovali primernost sintetičnih slik za učno množico. Pri tem smo se osredotočili tudi na velikost učne množice. To smo storili tako, da smo ustvarili podmnožice različnih velikosti in učili model zgolj na teh. Po učenju smo vedno preverili uspešnost naučenega modela na isti testni množici.

Učne množice so bile sestavljene iz 10, 20, 50, 100, 500, 1000 in 1500 slik. Pri resničnih slikah smo vzeli samo 10, 20, 50 in 100 slik, saj imamo le 200 slik v podatkovni množici, 100 pa jih je bilo namenjenih za vrednotenje. Zaradi odvisnosti rezultatov od specifičnih slik, ki so bile izbrane v majhne učne podmnožice, smo pri majhnemu številu učnih primerov (10–500) vsak eksperiment desetkrat ponovili, pri čemer smo vsakič izbrali druge učne primere. Kot lahko razberemo iz Slike 5, rezultati dobljeni na resničnih slikah še zmeraj presegajo rezultate dobljene na sintetičnih. Pri sintetičnih slikah ravno tako lahko opazimo, da obstaja optimalno število primerov, ki se nahaja okrog 500. Pri večjem številu sintetičnih učnih slik se namreč model preveč oz. še bolj prilagodi na specifične sintetične slike, zato se rezultati na pravih slikah poslabšajo. Pri tem je treba poudariti, da smo za določanje praga za segmentacijo uporabili kar v naprej določen prag (vrednost 0,5), ki ni vedno optimalna. Ob bolj primernem pragu bi se ti rezultati lahko popravili.

Tabela 1 prikazuje tudi rezultate v obliki povprečne natančnosti, ki ne zahteva vnaprejšnje nastavitve vrednosti praga. Na osnovi teh rezultatov lahko sklepamo, da dobimo uspešen model že pri manjšem številu sintetičnih učnih slik. Tipično pa v praksi potrebujemo končno binarno segmentacijsko sliko, zato je moramo določiti tudi prag. In ker v tem scenariju nimamo na razpolago resničnih slik za nastavitveno

množico, saj bi se radi zajemanju resničnih slik izognili, smo primorani fiksni prag določiti v naprej.



Slika 5: Presek nad unijo (pri pragu 0,5) v odvisnosti od velikosti učne množice

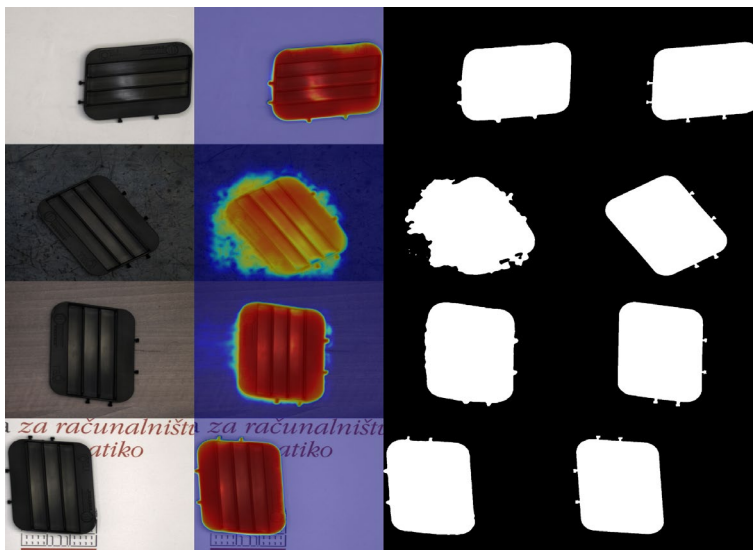
Vir: lasten.

Tabela 1: Rezultati prvega eksperimenta.

Št Učnih primerov	Tip slik	Povprečna natančnost	Presek nad unijo pri pragu 0,5
10	Resnične	99,56	82,33
10	Sintetične	98,93	62,75
50	Resnične	99,80	86,18
50	Sintetične	97,40	76,53
100	Resnične	99,94	96,98
100	Sintetične	98,05	79,97
500	Sintetične	98,17	83,34
1000	Sintetične	97,65	73,85
1500	Sintetične	96,39	78,99

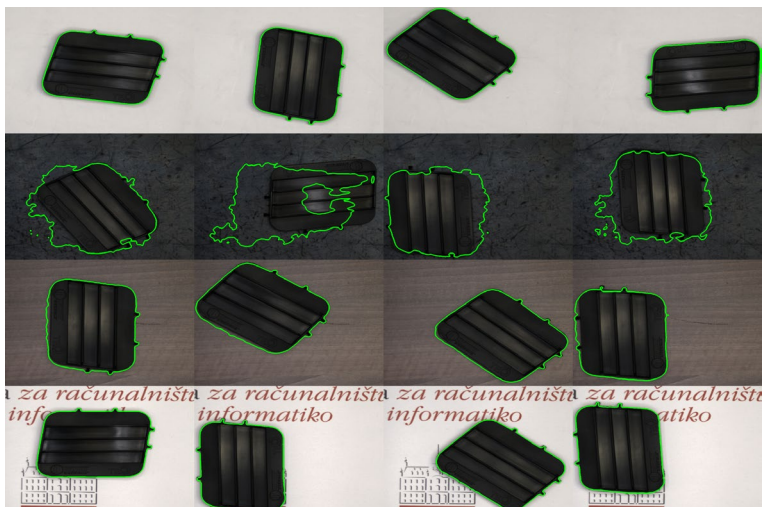
Na Slikah 6 in 7 lahko vidimo primer napovedi modela, ki se je učil na stotih sintetičnih slikah z naključnimi ozadji. Opazno je, da so rezultati zelo odvisni od ozadja testnih slik. Ne glede na število sintetičnih slik v učni množici so zaključki podobni. Na belem ozadju in na plakatu se model zelo lepo nauči tudi detajlov predmeta. Pri lesenem ozadju je kontrast med predmetom in ozadjem manjši, zato se ob učenju na sintetičnih slikah z naključni ozadjem nekaj detajlov na robu

predmeta izgubi. Najslabši rezultati so doseženi na slikah s črnim ozadjem, saj je kontrast med predmetom in ozadjem zelo majhen.



Slika 6: Primeri iz testne množice in napovedi modela pri učenju s 100 sintetičnimi slikami.

Vir: lasten.



Slika 7: Primeri napovedi modela, naučenega s stotimi sintetičnimi slikami, na različnih podlagah.

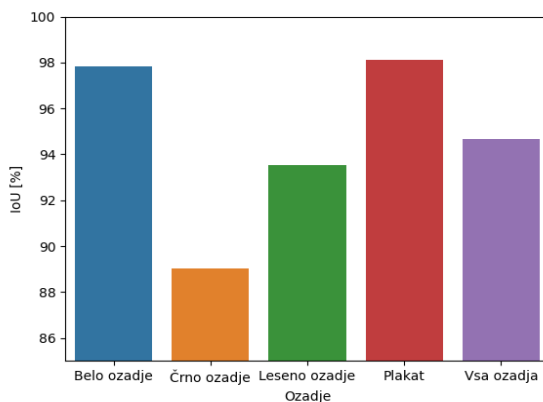
Vir: lasten.

5.2 Drugi eksperiment

Rezultati prvega eksperimenta torej kažejo na to, da je uspeh zelo odvisen od ozadja predmeta. Za pričakovati je, da se bodo rezultati bistveno izboljšali, če se model ne bo učil na naključnih ozadjih, temveč na ozadjih podobnim tistim v testni množici. To je tudi realen scenarij saj je približen izgled pravih ozadij v industrijskem okolju poznan.

V drugem eksperimentu smo tako analizirali primernost sintetičnih slik z vnaprej znanim ozadjem za uporabo segmentacij resničnih slik. Tako smo model učili na sintetičnih slikah z enim ozadjem in vrednotili na resničnih slikah posnetih na podobnem ozadju. Na koncu smo model naučili na vseh slikah z vsemi štirimi ozadji in ga tudi ovrednotili na pravih slikah z vsemi štirimi ozadji.

Iz Slike 8 in je mogoče razbrati, da učenje na sintetičnih slikah z istimi ozadji, kot pri resničnem zajemanju, prinese zadovoljive rezultate na resničnih slikah. Opazimo lahko, da je rezultat odvisen od težavnosti ozadja, saj dobimo boljše rezultate na belem ozadju in na plakatu, kjer nismo imeli problemov tudi v prvem eksperimentu.



Slika 8: Presek nad unijo (pri pragu 0,5) za različna ozadja.

Vir: lasten.

Tabela 2 prikazuje tudi rezultate v povprečni natančnosti. Na podlagi teh rezultatov lahko ponovno sklepamo, da je rezultat v veliki meri odvisen od ozadja, ki je bilo prisotno v učni množici.

Tabela 2: Rezultati drugega eksperimenta

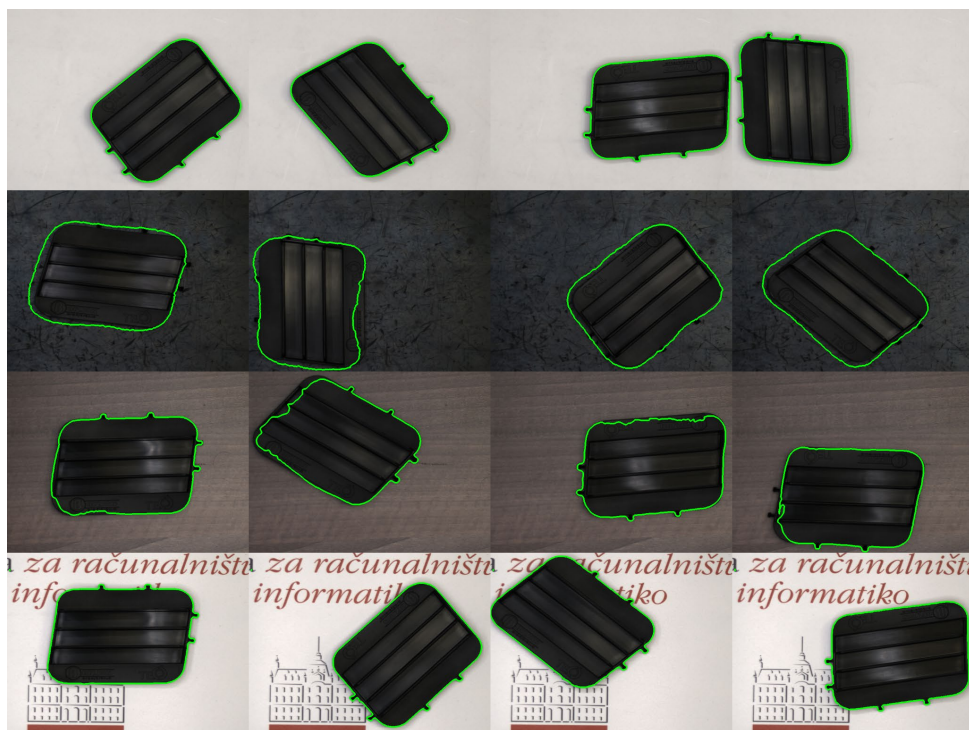
Št. učnih primerov	Ozadje	Povprečna natančnost	Presek nad unijo pri pragu 0,5
200	Belo	99,97	97,83
200	Črno	99,23	89,01
200	Leseno	99,64	93,53
200	Plakat	99,98	98,11
800	Vsa	99,71	94,67

Na Slikah 9 in 10 lahko vidimo nekaj primerov delovanja našega modela. Kot lahko razberemo se pri belem ozadju in plakatu lepo nauči tudi manjše detajle predmeta. Ravno tako se teh detajlov nauči pri lesenem ozadju, ampak nekoliko slabše. Najslabše deluje na črnem ozadju, kjer se nauči samo glavne oblike predmeta. Na kontrastnih prvem in zadnjem ozadju so rezultati torej blizu optimalnim, medtem ko so rezultati na manj kontrastnih ozadjih slabši.



Slika 9: Primeri iz testne množice in napovedi našega modela pri učenju modela na resničnih ozadjih.

Vir: lasten.



Slika 10: Primeri napovedi modela, naučenega s sintetičnimi slikami z enakimi podlagami kot pri resničnem zajemanju, na različnih podlagah.

Vir: lasten.

6 Zaključek

V tem članku smo predstavili cevovod za generiranje sintetičnih slik iz 3D CAD modelov. Predstavili smo zgenerirano množico slik in preverili uporabnost sintetičnih slik za učenje segmentacije resničnih slik. Pri pregledu uporabnosti teh slik smo proučili tudi vpliv velikosti učne množice slik. V skladu s pričakovanji smo ugotovili, da je model bolj uspešen, če mu podamo resnične slike. Pokazali smo tudi, da sintetična množica ne sme biti prevelika. Nato smo se osredotočili na problem, ko je ozadje vnaprej poznano. Pokazali smo na uporabnost sintetičnih slik s takim ozadjem, saj se rezultati modela naučenega zgolj na sintetičnih slikah približajo rezultatom modela dobljenega s pravimi slikami, dasiravno jih še ne dosežejo.

V prihodnjem delu nameravamo ta razkorak med rezultati dobljenimi s sintetičnimi in resničnimi slikami še zmanjšati. Tako nameravamo še izboljšati postopek generiranja sintetičnih slik, ga bolj obogatiti in povečati naključnost in različnost pojavnosti. Po drugi strani pa nameravamo učenje s sintetičnimi slikami nadgraditi z informacijo vsebovano v majhnem številu pravih slik. Tako bomo naslovili problem premika domene (ang. domain shift), ki se v industrijskem okolju pogosto zgodi tudi zaradi postopnih sprememb v okolju v katerem se izvaja industrijski proces. Raziskali bomo torej možnost adaptacije na domeno (ang. domain adaptation), tako preskok s sintetičnih na prave slike kot prilagoditev na spremembe pojavnosti zajetih slik, pri čemer bo glavno vodilo, da bodo te spremembe čim bolj preprosto izvedene, brez potrebe po zajemanju in predvsem označevanju novih pravih slik.

Opombe

To delo je bilo delno financirano s strani ARRS projektov L2-3169 (MV4.0) in J2-4457 (RTFM) ter raziskovalnega programa Računalniški vid (P2-0214). Posebna zahvala gre tudi podjetju TECOS in Alešu Hančiču za izdelavo predmetov in 3D CAD modela uporabljenega v eksperimentu.

Literatura

- Deng, J., Dong, W., Socher, R., Li, L., Li, K. and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. *2009 IEEE conference on computer vision and pattern recognition*, pages 248-25.
- Wang, Q., Gao, J., Lin, W., & Yuan, Y. (2019). Learning from synthetic data for crowd counting in the wild. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 8198-8207).
- Richter, S. R., Vineet, V., Roth, S., & Koltun, V. (2016). Playing for data: Ground truth from computer games. In *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14* (pp. 102-118). Springer International Publishing.
- McCormac, J., Handa, A., Leutenegger, S., & Davison, A. J. (2017). Scenetnet rgb-d: Can 5m synthetic images beat generic imagenet pre-training on indoor segmentation?. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 2678-2687).
- Pinaya, W. H., Tudosiu, P. D., Dafflon, J., Da Costa, P. F., Fernandez, V., Nachev, P., ... & Cardoso, M. J. (2022, October). Brain imaging generation with latent diffusion models. In *Deep Generative Models: Second MICCAI Workshop, DGM4MICCAI 2022, Held in Conjunction with MICCAI 2022, Singapore, September 22, 2022, Proceedings* (pp. 117-126). Cham: Springer Nature Switzerland.
- Akrouf, M., Gyepesi, B., Holló, P., Poór, A., Kincsó, B., Solis, S., ... & Fazekas, I. (2023). Diffusion-based Data Augmentation for Skin Disease Classification: Impact Across Original Medical Datasets to Fully Synthetic Images. *arXiv preprint arXiv:2301.04802*.
- Dewi, C., Chen, R. C., Liu, Y. T., & Tai, S. K. (2022). Synthetic Data generation using DCGAN for improved traffic sign recognition. *Neural Computing and Applications*, 34(24), 21465-21480.
- Mayer, N., Ilg, E., Haussler, P., Fischer, P., Cremers, D., Dosovitskiy, A., & Brox, T. (2016). A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4040-4048).

- Li, W., Pan, C. W., Zhang, R., Ren, J. P., Ma, Y. X., Fang, J., ... & Yang, R. G. (2019). AADS: Augmented autonomous driving simulation using data-driven algorithms. *Science robotics*, 4(28), eaaw0863.
- Huang, X., Cheng, X., Geng, Q., Cao, B., Zhou, D., Wang, P., ... & Yang, R. (2018). The apolloscape dataset for autonomous driving. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 954-960).
- Yogamani, S., Hughes, C., Horgan, J., Sistu, G., Varley, P., O'Dea, D., ... & Pérez, P. (2019). Woodscape: A multi-task, multi-camera fisheye dataset for autonomous driving. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 9308-9318).
- Zhan, W., Sun, L., Wang, D., Shi, H., Clause, A., Naumann, M., ... & Tomizuka, M. (2019). Interaction dataset: An international, adversarial and cooperative motion dataset in interactive driving scenarios with semantic maps. *arXiv preprint arXiv:1910.03088*.
- Mayershofer, C., Ge, T., & Fottner, J. (2021). Towards fully-synthetic training for industrial applications. In *LISS 2020: Proceedings of the 10th International Conference on Logistics, Informatics and Service Sciences* (pp. 765-782). Springer Singapore.
- Eversberg, L., & Lambrecht, J. (2021). Generating images with physics-based rendering for an industrial object detection task: Realism versus domain randomization. *Sensors*, 21(23), 7901.
- Abou Akar, C., Tekli, J., Jess, D., Khoury, M., Kamradt, M., & Guthe, M. (2022, October). Synthetic Object Recognition Dataset for Industries. In *2022 35th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)* (Vol. 1, pp. 150-155). IEEE.
- Quattrocchi, C., Di Mauro, D., Furnari, A., & Farinella, G. M. (2022, May). Panoptic Segmentation in Industrial Environments Using Synthetic and Real Data. In *Image Analysis and Processing—ICIAP 2022: 21st International Conference, Lecce, Italy, May 23–27, 2022, Proceedings, Part II* (pp. 275-286). Cham: Springer International Publishing.
- Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision* (pp. 2980-2988).
- Chen, L. C., Papandreou, G., Schroff, F., & Adam, H. (2017). Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*.
- Denninger, M., Sundermeyer, M., Winkelbauer, D., Zidan, Y., Olefir, D., Elbadrawy, M., ... & Katam, H. (2019). Blenderproc. *arXiv preprint arXiv:1911.01911*.
- Community, B. O. (2018). Blender - a 3D modelling and rendering package. Stichting Blender Foundation, Amsterdam. Retrieved from <http://www.blender.org>

NAPOVEDOVANJE OBRAZOV OTROK Z GAN METODO

NEJC ŠUKLJE, LUKA ŠAJN

Univerza v Ljubljani, Fakulteta za računalništvo in informatiko, Ljubljana, Slovenija
nejc.suklje@hotmail.com, luka.sajn@fri.uni-lj.si

Sinopsis Naslov "Napovedovanje slik otrok obrazov z GAN metodo" je že na prvi pogled vzbudil motivacijo, ker je potencialni rezultat zanimiv. GAN metoda je tudi dokaj novo področje, kar vse skupaj naredi še bolj zanimivo, saj ni veliko raziskav, v katerih je bila uporabljena. Na problem smo pogledali tudi s strani "uporabnika", torej nekoga, ki bi rad videl sliko svojega potencialnega otroka. Iz tega sklepamo, da bi končna rešitev na trgu doživela zanimanje, še posebej v bogatejšem sloju prebivalstva. Raziskav na to temo je bolj malo, vendar dovolj za osnovo pri izdelavi naloge. Cilj naloge je bil dobiti sliko obraza otroka na osnovi izgleda njegovih staršev.

Prispevek temelji na:

Šuklje, N. (2022).

Napovedovanje obrazov otrok z GAN metodo: diplomsko delo,

Univerza v Ljubljani, Fakulteta za računalništvo in informatiko. Ljubljana.

Ključne besede:

grebenska regresija, latentni prostor, generiranje obrazov, GAN metoda, sorodstvena podobnost

PREDICTING CHILDREN FACES WITH GAN METHOD

NEJC ŠUKLJE, LUKA ŠAJN

University of Ljubljana, Faculty of Computer Science and Informatics, Ljubljana,
Slovenia
nejc.suklje@hotmail.com, luka.sajn@fri.uni-lj.si

**The paper is
based on:**
Šuklje, N. (2022).
Napovedovanje
obrazov otrok z
GAN metodo:
bachelor's thesis,
University of
Ljubljana, Faculty
Computer and
Information
Science. Ljubljana.

Keywords:
ridge regression,
latent space,
GAN method,
generation of
faces,
kinship
resemblance

Abstract The title "Prediction of children's faces with the GAN method" aroused motivation at first glance, because the potential result is interesting. The GAN method is also a fairly new thing, which made it more interesting because there aren't many studies that have used it. We also looked at the problem from the "buyer's" side, that is, someone who would like to see a picture of their potential child. From this, we conclude that the final solution on the market would experience an increase, especially in the wealthier strata of the population. There is less research on this topic, but it still helped in the preparation of the assignment. The goal of the task was to get a picture of a child's face from its parents.

1 Uvod

GAN metoda je pritegnila veliko pozornosti med raziskovalci na področju računalniškega vida (Šuklje, 2022). Največ se uporablja v medicini za sintezo slik. Cilj je pridobiti čim večjo količino podatkov oziroma slik, ki so podobne realnim slikam. Te generirane slike so nato uporabljene skupaj z dejanskimi za potrebe učenja. Študenti medicine lahko na njih izvajajo analize in se tako učijo na novih izračunanih primerih. Zaradi raznolikosti, ki jo dodamo z namišljenimi slikami, je lahko pridobljeno znanje širše.

Delo je osnovano na že prej pripravljeni kodi Matjaža Mava (Šuklje, 2022). Osnova je zastavljena kot napovedovanje otrok na podlagi izgleda staršev z globokim učenjem (ang. deep learning). Začeli bomo s predstavitvijo GAN metode, potrebnimi splošnimi informacijami za razumevanje le-te, nato pa pridejo na vrsto postopki in metode, ki so bili uporabljeni, da smo prišli do želenega rezultata. Rezultati bodo predstavljeni in komentirani. Primerjana bo podobnost generiranih slik z resničnimi. Po analizi smo tudi podali možne ideje za nadaljnje delo.

2 Pregled področja

2.1 GAN metoda grobo

GAN metoda je relativno mlada metoda, ki se uporablja pri procesiranju slik (Šuklje, 2022). Uporablja se tako, da na začetku uporabimo učno množico slik, s katero se bo program učil obraznih značilnic. Ta množica se čez čas razširja, ker program proizvaja nove slike, ki so dodane v to množico in so približek tistim, ki so bile v množici že od začetka.

2.2 Prva omemba GAN metode

GAN metoda je bila prvič predstavljena in opisana v članku (Goodfellow in sod., 2014). Članek govori o modelu, ki je sestavljen iz dveh funkcij. Prva je zadolžena za generiranje podatkov, druga pa za oceno primernosti generiranih podatkov, in jih opredeli kot pravilne ali napačne. Cilj teh dveh funkcij je, da povečamo verjetnost prepoznave napake druge funkcije. S tem dosežemo, da je podatkov opredeljenih kot pravilnih malo, vendar so zelo podobni resničnim. Do te mere, da ne moremo ločiti med resničnimi in neresničnimi. Avtorji članka generator in diskriminator

razložita na zelo razumljiv način. Generator si lahko predstavljamo kot ponarejevalca, ki poskuša narediti čim bolj pristne ponaredke, ki bodo ostali neodkriti, diskriminator pa kot policijo, ki išče ponaredke in njihovo uporabo na trgu (Škuje, 2022). Modela nenehno izboljšujeta svoji funkciji (Goodfellow in sod., 2014).

2.3 Sorodna dela

Članek (Goodfellow in sod., 2014) je pritegnil veliko pozornosti med raziskovalci na področju računalniškega vida. Največ se uporablja v medicini za sintezo slik in posledično za pridobivanje večje količine podatkov (Yi in sod., 2019).

Članek (Rani, 2014) govori o generiranju slik otrok. Bolj specifično tistih, ki imajo avtizem. Zaradi bolezni je njihovo izražanje čustev različno od ljudi, ki nimajo avtizma. S procesiranjem njihovih izrazov na obrazu se poskuša določiti čustva, ki jih ta trenutek izražajo (Rani, 2014).

3 Kako deluje GAN metoda

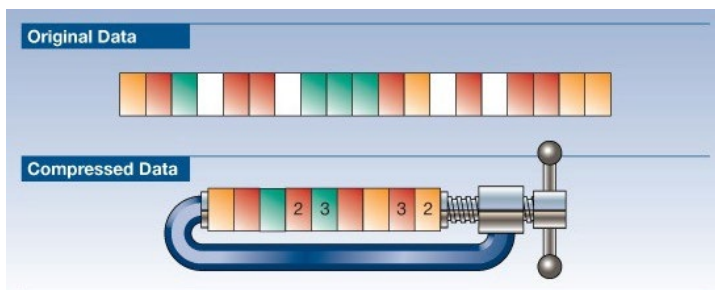
Kratice GAN pomeni "Generative Adversarial Network", v slovenščino to prevedemo kot generativna nasprotniška mreža. Metoda uporablja 2 nevronske mreže, ki delujeta ena proti drugi. Iz tega tudi izhaja beseda v imenu metode, in sicer "Adversarial". Nevronske mreže ustvarjata podatke, ki bi lahko veljali za resnične (Goodfellow in sod., 2014).

Za boljše razumevanje metode bomo navedli konkreten primer. Opisali bomo tudi generator in diskriminator, ki sta ključni funkciji v metodi (Škuje, 2022). Predpostavimo, da generativni algoritem generira nam nepoznane podatke, za katere ne poznamo njihovih značilnosti. Zdaj je naloga diskriminatorja, da nam te podatke kategorizira in posledično označi. Glede na značilnosti podatkov dobimo oznako. Vhodne podatke seveda po potrebi razdelimo na več vrst. Kot primer vzemimo e-pošto. Naloga diskriminatorja je, da nam označi posamezno sporočilo kot nezaželeno/vsiljeno ali pa običajno. Diskriminator vhodne podatke razdeli na besede v sporočilu in na naslov sporočila (zadeva). Matematično to lahko predstavimo s pomočjo verjetnostnega zapisa. Oznako označimo z z , kot značilnost oziroma z x , kot podatke. Formula, ki povezuje ti dve neznanki je $p(y | x)$. Formulo preberemo kot "verjetnost y pri x ", v našem konkretnem primeru bi se to glasilo kot "verjetnost, da je sporočilo nezaželeno glede na besede v sporočilu".

Torej diskriminator glede na značilnosti podatkov le-te označi. Zdaj se vrnemo k generatorju. Njega si lahko predstavljamo, da deluje obratno. Torej, da iz označbe dobi značilnosti. To pomeni, da bo zgeneriral podatke z določenimi značilnostmi, ki veljajo za neko podano označbo. Imamo še en način, kako lahko opišemo generator in diskriminator. Diskriminator pozna meje med razredi, generator pa pozna porazdelitev vsakega razreda posebej (Nicholson, 2020).

4 Latentni prostor

Latentni prostor je izraz, ki ga moremo razumeti, da nam bo celotna slika bolj jasna. V našem programu smo operirali s slikami in to na tak način, da smo jih predstavili vektorsko, torej s številkami. Problem, ki se tu pojavi, je količina podatkov, ki jih posamezna slika vsebuje. Zmanjšati moramo dimenzije, ki jih porabimo za predstavitev slike. Dimenzija novo nastalega vektorja mora biti občutno manjša. Vektor z zmanjšanimi dimenzijami, ki ga dobimo, se imenuje latentna predstavitev slike. Na sliki 1 je predstavljena ideja manjšanja dimenzij. Devetnajst točk stisnemo tako, da jih lahko prestavimo samo z devetimi (Manning, 2022), (Silva, 2018), (Tiu, 2020).



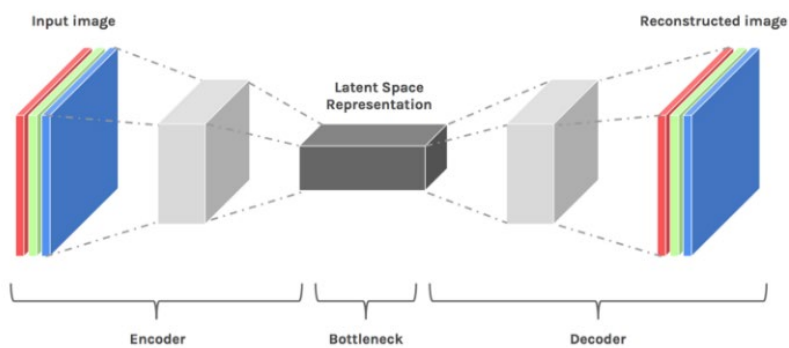
Slika 1: Primer kompresije vektorja.

Vir: (Manning, 2022)

Torej imamo za vsako sliko, ki jo bomo uporabili v našem generatorju, latentno predstavitev v obliki vektorja (Šuklje, 2022). Vsi ti vektorji sestavljajo oziroma oblikujejo prostor, ki mu rečemo latenten prostor. V tem prostoru so slike, torej njihovi vektorji, ki so si podobni bolj skupaj, slike, ki si pa niso, pa so bolj oddaljene. Dimenzija latentnega prostora je določena z velikostjo vektorja. Na primer, če imajo vektorji dolžino 3, to pomeni, da je dimenzija latentnega prostora 3. Na sliki 2 imamo prikazano vizualizacijo latentnega prostora. Slike, ki jih stisnemo, sestavljajo latenten

prostor tako, da so ohranjeni samo pomembni podatki. To je tudi osnova, da sploh lahko stiskamo. Seveda, da lahko iz latentnega prostora dekompresiramo kakšno sliko, potrebujemo dekoder, ki je to sposoben narediti (Manning, 2022), (Silva, 2018), (Tiu, 2020).

Ime »latent space« oziroma latentni prostor izhaja iz besede »hidden« oziroma skrit. To pa zato, ker si težko predstavljamo prostor, ki ima kakšno dimenzijo več kot 3, kaj šele prostore dimenzije reda 100 (Manning, 2022), (Silva, 2018), (Tiu, 2020).



Slika 2: Vizualizacija latentnega prostora.

Vir: (Manning, 2022)

5 Rezultati delovanja programa

Ko je bil program enkrat razvit smo potrebovali slike, da program uporabimo na dejanskem primeru. Za učenje smo uporabili 217 primerov. Torej 1 primer pomeni: slika očeta, slika mame in slika njunega otroka. Za testiranje pa smo imeli 20 takih primerov. Vse slike, ki smo jih uporabili, smo dobili že iz vnaprej pripravljene baze (Silva, 2018). Njihove latentne slike pa nam je zagotovil laboratorij za računalniški vid na fakulteti za računalništvo in informatiko. Napovedovalni model smo zgradili z grebensko regresijo. Testne slike smo dali v model in zadnji korak je bil le, da še te latentne slike generiramo in dobimo dejanske slike. To nam je omogočila grafična kartica NVIDIA GeForce GTX 970M. Hitrost je 924 MHz, velikost spomina pa 3GB. Generiranje 80 slik je trajalo nekaj manj kot 5 minut. Slike smo analizirali sami, za dejansko vrednotenje kakovosti smo uporabili ocene zunanjih neodvisnih ocenjevalcev, ki so podali svoje mnenje. S tem smo lažje zaključili in ocenili celotno delo.

5.1 Analiza

Na naslednji sliki, sliki 3, imamo v vsaki vrstici o leve proti desni: očeta, mamo, njunega otroka in na koncu še generiranega otroka z našim programom. Vsaka vrstica je svoj primer, torej imamo 4 primere.



Slika 3: Primeri staršev, njunega otroka in generiranega otroka

Vir: lasten

To so 4 primeri slik, ki smo jih pokazali zunanjim opazovalcem. Dobili smo mnenja ljudi od starosti 20 pa vse do 50 let. Ob prebiranju mnenj smo ugotovili, da imajo različne starostne skupine drugačno razmišljanje in, da drugače opazajo podrobnosti in celostno sliko. Mlajša starostna skupina, torej okrog 20 let, ni opazila manjših podrobnosti, so pa bila podana mnenja zelo razumljiva in smiselno sestavljena. Starejša starostna skupina, 40-50 let, pa je pokazala čisto nasprotje. Opazili so veliko manjših podrobnosti, vendar so bila njihova mnenja napisana manj enotno, večasih zmedena.

5.2 Naše mnenje

Nad podobnostjo resničnih in generiranih otrok smo bili kar malo presenečeni saj nismo pričakovali, da je možno doseči takšno podobnost. Glede na to, da so generirani otroci zgrajeni le na podlagi izgleda staršev, ne pa tudi genskega zapisa, so v večini bolj podobni samim staršem kot njihovi resnični otroci. Ne najbolj optimalna rešitev, vendar za začetek deluje dovolj dobro.

Na prvi pogled padeta v oči prva in tretja vrstica. Resničen in generiran otrok sta si tako zelo podobna, da smo prepričani, da bi ju večina opazovalcev videla kot sorojenca. Pri prvi in četrti vrstici sta si generirana otroka zelo podobna. Najverjetneje tukaj igra vlogo to, da imata podobni materi. Drugih specifičnih podobnosti ne opazimo, zato bomo zdaj podali še splošno mnenje.

Na prvi pogled zgleda, kot da je generiran otrok podoben staršu istega spola, vendar že pri prej analiziranem primeru, ko smo rekli, da imata primerka podobni materi, se zdi, da to ni čisto res. Predvidevamo, da nas k temu zavede dolžina las in zato otrok izgleda podoben staršu istega spola. Zdi se pa tudi, da so generirani otroci zelo podobni resničnim otrokom, čeprav ti niso bili uporabljeni pri generiranju. S tem dobimo potrditev, da je resničen otrok podedoval veliko videza od staršev, najverjetneje več kot od starih staršev, ki niso uporabljeni v modelu.

6 Zaključek

S končnimi rezultati smo zadovoljni, ker je podobnost smiselna. Dobre rezultate potrjujejo tudi mnenja zunanjih opazovalcev. Izprašani dajejo občutek, da so zelo presenečeni nad podobnostjo. Poleg mnenj v pisni obliki smo z njimi tudi govorili, tam pa je bila presenečenost še bolj izražena. Predvidevamo, da je presenečenost posledica tega, da nimajo predznanj o tej temi in se jim zdi vsaka podobnost izjemna.

6.1 Možne izboljšave

Za nadaljnje delo bi predlagali, da se za generacijo otrok ne uporabi enak delež očeta in matere. Predvidevamo, da bi takšni rezultati bili še bolj zanimivi, ker bi bil otrok moškega spola še bolj podoben očetu in bi lahko iskali po obrazu, kaj je pa tisto, kar je podedoval po mami. Glede na to, da je generiranje temeljilo le na izgledu staršev

bi bilo smiselno v generiranje zraven staršev vključiti tudi stare starše, vendar seveda z manjšim deležem.

Literatura

- Goodfellow I.J. in sod. (2014). *Generative Adversarial Networks*. doi: 10.48550/ARXIV.1406.2661. url: <https://arxiv.org/abs/1406.2661>
- Yi X., Walia E. in Babyn P. (2019). *Generative adversarial network in medical imaging: A review*. V: *Medical Image Analysis* 58, str. 101552. issn: 1361-8415. doi: <https://doi.org/10.1016/j.media.2019.101552>. url: <https://www.sciencedirect.com/science/article/pii/S1361841518308430>.
- Rani P. (2014). *Emotion Detection of Autistic Children Using Image Processing*. V: *2019 Fifth International Conference on Image Information Processing (ICIIP)*. Nov. 2019, str. 532–535. doi: 10.1109/ICIIP47207.2019.8985706
- Nicholson C. (2020). *A Beginner's Guide to Generative Adversarial Networks (GANs)*. url: <https://wiki.pathmind.com/generative-adversarial-network-gan> (pridobljeno 27. 10. 2022)
- Manning publications. *concept latent space in category GAN*. (2022). url: <https://livebook.manning.com/concept/gan/latent-space> (pridobljeno 29. 10. 2022)
- Silva T. (2018). *An intuitive introduction to Generative Adversarial Networks (GANs)*. url: <https://www.freecodecamp.org/news/an-intuitive-introduction-to-generative-adversarial-networks-gans-7a2264a81394> (pridobljeno 29. 10. 2022)
- Šuklje, N. (2022). *Napovedovanje obrazov otrok z GAN metodo*: diplomsko delo, Univerza v Ljubljani, Fakulteta za računalništvo in informatiko. Ljubljana.
- Tiu E. (2020). *Understanding Latent Space in Machine Learning*. url: <https://towardsdatascience.com/understanding-latent-space-in-machine-learning-de5a7c687d8d> (pridobljeno 29. 10. 2022)
- Robinson J.P. in sod. (2021.) *Families in Wild Multimedia: A Multimodal Database for Recognizing Kinship*. V: *IEEE Transactions on Multimedia* 24, str. 3582–3594
- Šuklje N. (2022). *Napovedovanje obrazov otrok z GAN metodo*. Univerza v Ljubljani, Diplomsko delo.

DETEKCIJA PUŠČIC PRI KLASIČNEM PIKADU

MATIC ZGONC, BORUT BATAGELJ

Univerza v Ljubljani, Fakulteta za računalništvo in informatiko, Ljubljana, Slovenija
mz4144@student.uni-lj.si, borut.batagelj@fri.uni-lj.si

Sinopsis V članku je opisan razvoj sistema za detekcijo puščic pri klasičnem pikadu, z uporabo metod računalniškega vida, ki so implementirane v odprtokodni knjižnici OpenCV. Cilj je bil razviti sistem z visoko natančnostjo, ki bo enostaven za uporabo in cenovno dostopen. Na začetku je opisan postopek kalibracije tarče, sledi iskanje točke udarca in izračun točk in na koncu še uporabljene prilagoditve pri razvoju sistema.

Ključne besede:

OpenCV,
računalniški vid,
pikado,
homografija,
kalibracija

ARROW DETECTION FOR CLASSIC DARTS

MATIC ZGONC, BORUT BATAGELJ

University of Ljubljana, Faculty of Computer Science and Informatics, Ljubljana,
Slovenia
mz4144@student.uni-lj.si, borut.batagelj@fri.uni-lj.si

Abstract This paper describes the development of a system for arrow detection in classic darts using computer vision methods in the open-source library OpenCV. The goal was to develop a highly accurate system that is easy to use and affordable. At the beginning, the procedure of target calibration is described, followed by the search for the hit point and the calculation of the score, and finally a description of the adjustments used in the development of the system.

Keywords:
OpenCV,
computer vision,
darts,
homography,
calibration

1 Uvod

Pikado je šport, ki je pri nas in po svetu vedno bolj priljubljen. V zadnjih letih se opaža velika rast profesionalnih igralcev, gledalcev in denarnih nagrad. Na profesionalnih tekmovanjih se največkrat uporablja klasična, pri amaterskih igralcih pa elektronska tarča. Najpogostejši razlog je, da elektronska tarča omogoča štetje rezultata in podpira različne vrste iger, medtem, ko moramo pri klasični vse štetje opraviti ročno. To nas je vodilo k razvoju sistema, za avtomatsko zaznavanje puščic in štetje točk.

Na trgu že obstajajo rešitve, ki rešujejo ta problem. Najbolj znana in razširjena v naši okolici je Scolia (Scolia, 2023). Deluje na principu uporabe več kamer iz različnih zornih kotov, ki služijo zaznavanju puščic. Njegova glavna slabost je ročna kalibracija sistema, ki jo je potrebno ponoviti ob vsaki spremembi stanja tarče oz. pozicije kamer. Pri izdelavi našega sistema, smo si zadali implementiranje avtomatske kalibracije, ki jo lahko ponovimo kadarkoli med njegovim delovanjem.

Izgradnjo sistema smo razdelili na 3 sklope, in sicer kalibracijo tarče, iskanje točke udarca puščice, ter pridobivanje rezultata posameznega meta.

2 Izgradnja sistema

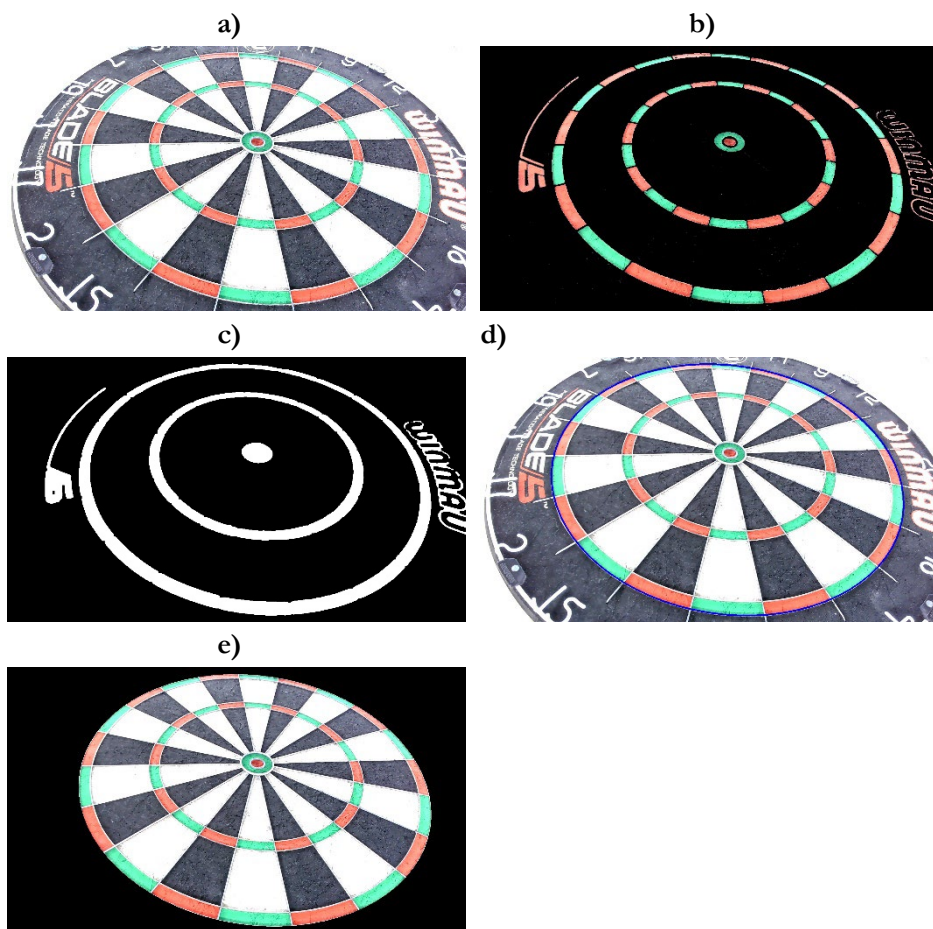
2.1 Kalibracija tarče

Pri kalibraciji tarče je cilj pridobitev transformacijske matrike. Ta nam vhodno sliko, na kateri je tarča v obliki elipse, pretvori v pravilno okroglo obliko. Uporabimo jo tudi pri pridobivanju točke udarca in rezultata meta.

Transformacijsko matriko pridobimo iz 4 točk, ki se na igralni površini tarče nahajajo 90° ena od druge. Pri pridobivanju transformacijskih točk, smo si pomagali s HSV (hue, saturation, value) barvnim prostorom, ter morfološki operacijami (Chhikara, 2022; Soille, 1999).

Prvi korak je iskanje elipse, ki obkroža igralno površino. Pridobimo jo z uporabo dveh HSV mask, eno za zeleno barvo in drugo za rdečo, ki jih združimo. Posamezna maska, na sliki ohrani slikovne točke, ki so v njenih mejah, ostale zavrže in jih označi z 0. Tako pridobimo dva kolobarja z zelenimi in rdečimi polji, ki pa med seboj niso

povezani. Z uporabo morfoloških transformacij dilacije in erozije pridobimo kolobar, ki je primeren za iskanje obrisov (angl. contours). Pridobljen obris že kaže podobo elipse, ki pa ni čisto pravilne oblike in vsebuje manjše nepravilnosti. Funkcija *fitEllipse* iz OpenCV paketa (Laganiere, 2017), po metodi najmanjših kvadratov izračuna elipso, ki se najbolj prilega na dan nabor 2D točk. Z izračunano elipso na koncu iz vhodne slike izločimo zunanje območje elipse. S tem odstranimo morebitne motnje, ki bi lahko oteževale nadaljnjo kalibracijo. Koraki do pridobitve elipse, so prikazani na Sliki 1.



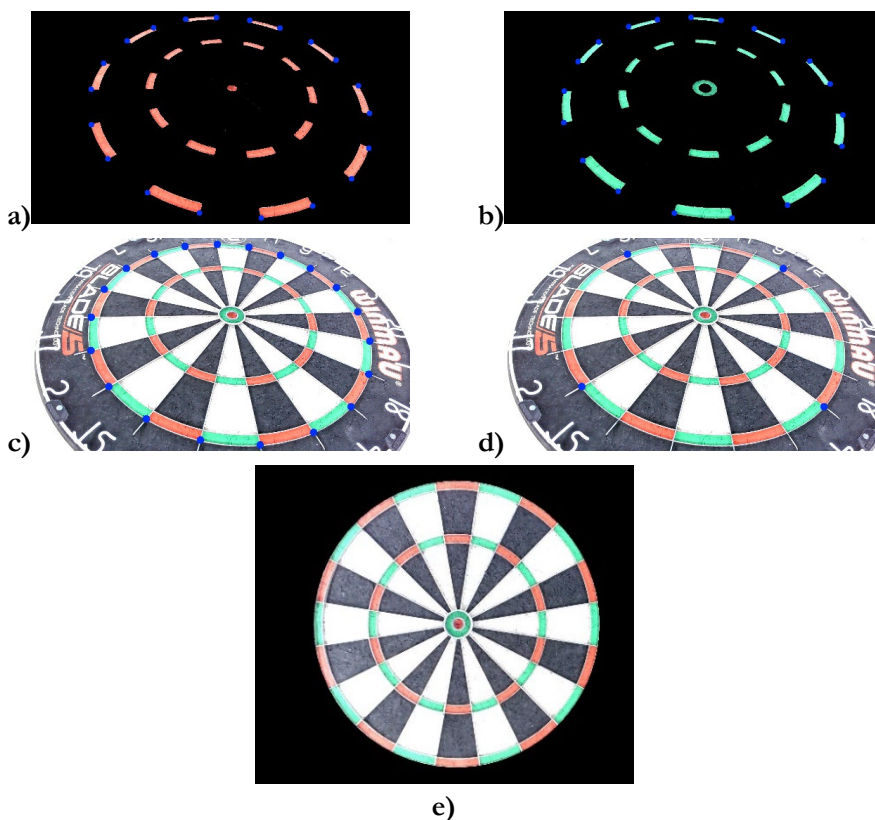
Slika 1: Vhodna slika (a), HSV maski (b), morfološke operacije (c), poiskana elipsa (d), izrezana slika tarče iz okolice (e)

Vir: lasten.

Naslednji korak je iskanje točk, ki razmejujejo sektorje na zunanjem krogu. Ponovno smo uporabili HSV maske, vendar tokrat ločeno. Tako pridobimo točke, ki razmejujejo zelene in rdeče sektorje. Za vhod v transformacijsko matriko izberemo štiri točke, ki so med seboj oddaljene 5 sektorjev.

Pred izvedbo transformacije, izračunamo še homografsko matriko (angl. Homography) z uporabo funkcije *findHomography*, ki mapira točke med dvema ravninama.

Zadnji korak je izvedba transformacije z uporabo funkcije *warpPerspective*. Ta na vhod dobi homografsko matriko, vhodne točke, ter ciljne točke, ki predstavljajo točke na poravnani tarči. Rezultat transformacije je kalibrirana tarča. Slika 2 prikazuje iskanje točk in kalibrirano sliko.



Slika 2: Posamezni HSV maski (a, b), točke zunanjih sektorjev (c), izbrane točke za transformacijo (d), kalibrirana slika (e)

Vir: lasten.

2.2 Iskanje točke udarca

Iskanje puščice smo izvedli s principom absolutne razlike med dvema slikama. Prva je posneta pred metom, druga pa po tem, ko puščica prileti v tarčo. Če so svetlobni pogoji v obeh primerih podobni, nam po izračunu razlike, ostane na sliki le puščica brez okolice. Da lahko poiščemo točko udarca, sliko najprej binariziramo, ter nato poiščemo obris puščice. Zaradi oblike puščice, točka udarca predstavlja najnižje ležečo točko na njej. Dobljene koordinate preslikamo v kalibrirano ravnino, da lahko izračunamo rezultat meta. Slika 3 prikazuje korake od izračuna absolutne razlike, do preslikane točke udarca.



Slika 3: Koraki do pridobitve točke udarca.

Vir: lasten.

2.3 Izračun točk

Tarča je razdeljena na 20 sektorjev, kar pomeni, da je vsak širok 18° . Z izračunanim kotom med preslikano točko udarca in središčem kalibrirane tarče, dobimo vrednost zadetih točk. Z uporabo kota lahko pridobimo le točke od 1 do 20, ne pa tudi vrednosti sredinskih krogov, ki sta vredna 25 in 50 točk, ter dvojnega in trojnega multiplikatorja. Ker so dimenzije tarč standardizirane, lahko za izračun teh vrednosti uporabimo razdaljo od točke udarca, do središča tarče. Za izračun končnih točk, enostavno pomnožimo vrednost zadetih točk z multiplikatorjem.

3 Prilagajanje sistema

V procesu razvoja smo izvedli različne prilagoditve, da bi izboljšali delovanje sistema in dosegali čim boljšo uspešnost zaznavanja. Primerjali smo različne osvetlitve tarče, različne kote namestitve kamere, uporabo kamer različnih kakovosti in namestitev kamere na različnih lokacijah. Pri vsaki prilagoditvi smo izdelali testne množice slik (Tabela 1), in na podlagi tega izbrali ustrezno rešitev, ki prinaša najboljše rezultate.

Tabela 1: Specifikacija testnih množic

Prilagoditev	Velikost testne množice
Osvetlitev tarče	111
Kot namestitve kamere	120
Različni kameri	117
Različne postavitve kamere	123

Pri osvetlitvi tarče smo preizkušali uspešnost zaznavanja v naravni osvetlitvi prostora in z uporabo led trakov. Za ta namen smo izdelali leseno ohišje za tarčo, okoli katerega smo namestili osvetlitev. Uporabili smo dve konfiguraciji led trakov. V prvi smo uporabili trak srednje moči, naravne barve svetlobe, v drugi pa smo poleg dodali še močnejši trak hladne bele svetlobe.

Ugotovili smo, da je za zaznavanje najboljša uporaba katerekoli konfiguracije led trakov, saj so bile vse puščice dobro vidne in pravilno zaznane, medtem, ko pri naravni osvetlitvi več kot 15% puščic sistem ni zaznal. Največ težav je bilo zaradi senc in pri puščicah v črnih poljih. Uporaba led trakov je te težave odpravila.

Pri naslednji prilagoditvi smo iskali kot, pod katerim kamera najboljše zaznava puščice. Identične kamere smo namestili na isto os, ki je bila pravokotna na tarčo. Preizkusili smo tri kote, in sicer 40° , 52° , in 60° . Izkazalo se je, da v uporabljeni testni

množici, sistem najboljšje zaznava puščice pri kotu 40° , kjer je bilo 5 prekritih puščic. Na ostalih dveh pozicijah, je prišlo do 8 primerov, kjer je peresce prekrilo konico druge puščice in onemogočilo zaznavo. Z večanjem kota, se povečuje zakritost vidnega polja zaradi peresca.

Preverili smo tudi, če kamera različne kakovosti in cenovnega razreda, vpliva na uspešnost. Izbrali smo spletni kameri, z ločljivostjo vsaj 1080p, saj dosegajo najboljše razmerje med kvaliteto in ceno. Prva je znamke Logitech, s senzorjem velikosti 2MP, ki omogoča zajemanje slik do ločljivosti 1920x1080 slikovnih točk in stane okoli 50€. Druga je znamke EKACOM, s 5MP senzorjem in najvišjo ločljivostjo 2560x1440 slikovnih točk. Njena cena je bila v času nakupa dobrih 25€. Pri testiranju smo izklopili samodejno ostrenje, ostale nastavitve pa so ostale privzete.

Uspešnost na testni množici je bila pri kameri EKACOM 98 %, pri Logitech pa 96 %. Menimo, da bi z manjšimi popravki nastavitvev, tudi pri Logitech kameri dosegli podoben rezultat. Tako smo pri razvoju sistema uporabili kamero EKACOM, saj je cenovno ugodnejša, pri isti uspešnosti kot dražja alternativa.

Nazadnje smo preverili, če postavitve kamere na različnih lokacijah okoli tarče vpliva na uspešnost. Izbrali smo postavitev na levi, desni in zgornji strani tarče. Na vse lokacije smo postavili identične kamere, pod istim kotom glede na tarčo. Ugotovili smo, da lokacija postavitve kamere ne vpliva na uspešnost zaznavanja. Na vseh lokacijah se pojavlja problem prekrivanja, njihovo število je naključno, odvisno od pozicije puščic. Dosežena uspešnost na poziciji desno in levo od tarče je bila slabih 97 %, iz zgornje strani pa 96%.

Edini problem, ki je ostal po vseh prilagoditvah je prekrivanje puščic. Enostavno in učinkovito rešitev za to težavo, smo dobili pri testiranju zadnje prilagoditve. Čeprav se je na vsaki izmed lokacij pojavljalo prekrivanje, smo ugotovili, da problemi niso bili prisotni na istih primerih. Izkazalo se je, da sta v vsaki situaciji, vsaj dve kameri vrnili točno vrednost. Eden izmed razlogov so različne lokacije kamer. Ker je razmak med njimi približno 120° , obstaja zelo majhna možnost, da bi bila ena puščica zakrita iz dveh lokacij. V primeru, ko vsaj dve kameri vrneta isto vrednost, privzamemo, da je to pravilen rezultat. S tem pristopom smo na isti testni množici dosegli uspešnost 100 %. Slika 4 prikazuje, končno konfiguracijo sistema s tremi kamerami.



Slika 4: Konfiguracija sistema po vseh prilagoditvah.

Vir: lasten.

4 Zaključek

V članku smo opisali razvoj in potrebne prilagoditve sistema za detekcijo puščic pri klasičnem pikadu. Ugotovili smo, da zaradi težav s prekrivanjem ena kamera ne zadošča za doseganje visoke uspešnosti. V končni različici sistema smo uporabili tri kamere in na testni množici dosegli 100 % uspešnost.

V prihodnje načrtujemo izdelavo sistema, še z uporabo metod računalniškega vida, ki temeljijo na strojnem učenju. Menimo, da bi s tem pristopom dobili bolj robusten sistem, ki bi bil bolj prilagodljiv na spremembe okolice.

Literatura

- Laganiere, R. (2017). OpenCV 3 Computer Vision Application Programming Cookbook, tretja izdaja. Packt Publishing Limited.
- Scolia. (2023) <https://scoliadarts.com/>.
- Chhikara, P. (2022). Understanding morphological image processing and Its operations. <https://towardsdatascience.com/understanding-morphological-image-processing-and-its-operations-7bcf1ed11756>.
- Soille, P. (1999). Morphological Image Analysis, Principles and Applications. (Vol. 2, No. 3, pp. 170-171). Berlin: Springer.

UNEXPLODED ORDNANCE DETECTION ON UAV THERMAL IMAGES BY USING YOLOv7

MILAN BAJIĆ,¹ BOŽIDAR POTOČNIK²

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

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

Abstract A few promising solutions for thermal imaging Unexploded Ordnance (UXO) detection were proposed after the start of the military conflict in Ukraine in 2014. Our research focuses on improving the accuracy of UXO detection in thermal images. The current state-of-the-art UXO detection method is based on the YOLOv5 Convolutional Neural Network (CNN). We assessed the effectiveness of UXO detection by using the state-of-the-art object detector YOLOv7 in this article. Two YOLOv7 models were re-implemented, fine-tuned using a grid-search approach and trained on a UXOTi_NPA public dataset of 720x480 pixel thermal images. The results showed that the models were able to identify UXOs from 11 different classes with more than 90% probability and a Mean Average Precision (mAP) of 86.8% to 89.7%, depending on the model's complexity. The metrics are just slightly behind the YOLOv5 results. Such CNN, thus, enables accurate automatic UXO detection, which is crucial to address one of the least explored and life-threatening problems worldwide.

Keywords:
unmanned aerial
vehicle;
unexploded
ordnance; thermal
imaging;
UXOTi_NPA
dataset;
convolutional
neural networks;
deep learning;
YOLO

DETEKCIJA NEEKSPLODIRANIH UBOJNIH SREDSTEV NA TERMALNIH SLIKAH UAV S POMOČJO YOLOV7

MILAN BAJIĆ,¹ BOŽIDAR POTOČNIK²

¹ Zagrebska univerza uporabnih znanosti, Oddelek za informatiko in računalništvo,
zagreb, Hrvaška
mbajic@tvz.hr

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

Sinopsis Po začetku vojaškega spopada v Ukrajini leta 2014 je bilo predlaganih nekaj obetavnih rešitev za odkrivanje neeksplodiranih ubojnih sredstev (UXO) v termalnih slikah. Naša raziskava se osredotoča na izboljšanje natančnosti detektiranja UXO v termalnih slikah. Trenutno najsodobnejša metoda detektiranja UXO temelji na konvolucijski nevronske mreži (CNN) YOLOv5. V tem članku smo ovrednotili učinkovitost detektiranja UXO s pomočjo najsodobnejšega detektorja objektov YOLOv7. Reimplementirali smo dva modela YOLOv7, ju fino uglasili s pristopom iskanja po mreži in ju naučili na javni zbirki UXOTi_NPA, ki vsebuje termalne slike velikost 720x480 pikslov. Rezultati so pokazali, da sta modela zmožna identificirati UXO iz 11 različnih razredov z več kot 90-odstotno verjetnostjo in s srednjo povprečno točnostjo (mAP) od 86,8 % do 89,7 %, odvisno od kompleksnosti modela. Rezultati le malo zaostajajo za rezultati YOLOv5. Takšna CNN torej omogoča natančno avtomatsko detekcijo UXO, kar je ključnega pomena za reševanje enega najmanj raziskanih in življenjsko nevarnih problemov na svetu.

Ključne besede:

EU projekt
AIM@VET,
računalniški vid,
umetna inteligenca,
poklicno
usposabljanje,
izobraževanje

1 Introduction

Thermal imaging is a non-contact method that converts the radiation pattern of an object into a visible image, known as a thermal image or thermogram. All objects above absolute zero temperature emit infrared radiation and this energy is detected by air and space-borne Thermal Infrared (TIR) remote sensors (Kuenzer & Dech, 2013). These systems do not require an external source of infrared radiation, making it suitable for field use.

Explosive remnants of war have been part of daily life since World War I, with 64 countries (Roberts & Williams, 1995) contaminated by landmines. There are countries with databases of explosive devices (GICHD, 2022), while others are still developing such information. Mine clearance is a slow process, taking approximately 100 times more time to clear each mine placed. Most of these remnants are well known by their physical dimensions, weight, type of fuse, and material of the cover. They mostly lie on the surface and are unaffected by weather or vegetation.

In a project by Norwegian People's Aid (NPA) in Bosnia and Herzegovina in 2019, thermal imaging (Bajic, 2020) was tested as a promising technology for surveying hazardous areas. In this research, thermal imaging and LiDAR sensors were mounted on Unmanned Aerial Vehicles (UAVs) and experimented with for this purpose.

Object detection by using thermal images is a broad research field (Leira et al., 2021), (Dai et al., 2021), (Banuls et al., 2020), however the detection of Unexploded Ordnances (UXOs) using thermal imaging is a relatively new research field. We found just few studies on this topic, like (Nikulin et al., 2018) for the PFM-1 'Butterfly Mines' detection and study (Krause et al., 2018) focused on the detection of landmines based on thermal changes in the environment. An experiment in (Yao et al., 2019) was conducted using time series thermal images to detect buried landmines. The use of Deep Learning for UXO detection is still relatively rare in the field. A research group from Binghamton University expanded their previous work (Nikulin et al., 2018) by using Deep Learning for automated detection and mapping of PFM-1 mines (Baur et al., 2020).

Recently, in our previous work (Bajić & Potočnik, 2023) we proposed the first automated Convolutional Neural Network (CNN) based solution for UXO detection using thermal imaging data that can identify multiple classes of objects. The YOLOv5 architecture was adapted successfully for the UXO detection problem in our previous study. Besides, the UXOTi_NPA public database (Bajić & Potočnik, 2023) of 808 annotated thermal images of Unexploded Ordnance was published. YOLOv5 model detection results on the UXOTi_NPA dataset are considered as the baseline results for this dataset.

In this study, we will assess the effectiveness of the newest YOLO architecture, i.e. YOLOv7, on the UXOTi_NPA public dataset. The YOLOv7 architecture will be adapted and retrained to detect UXOs from 11 different classes in highly variable thermal images. This architecture will be modified to fit the UXO detection problem, fine-tuned with a grid-search approach, and finally trained end-to-end on thermal images. The effectiveness of the adapted and retrained CNN architecture will be confirmed through experiments on the UXOTi_NPA dataset.

2 Evaluation Dataset UXOTi_NPA

The construction of the evaluation dataset UXOTi_NPA is described in (Bajić & Potočnik, 2023). This dataset consists of 808 thermal images of the dimension of 720×480 pixels with annotations. The number of objects (UXOs) per image varied between one to three. Figure 1 depicts a sample thermal image from our dataset.

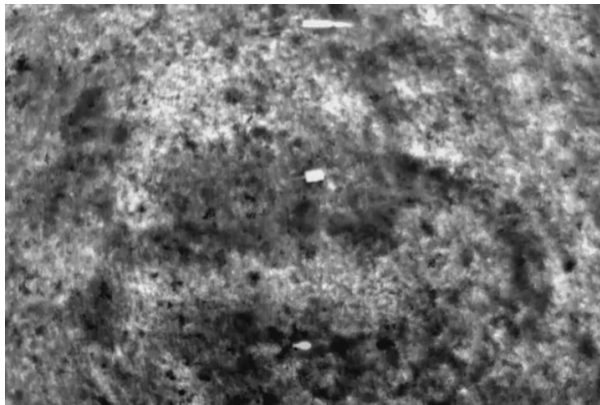



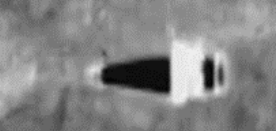

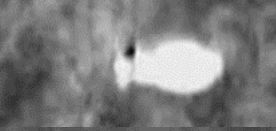







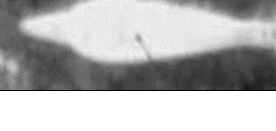







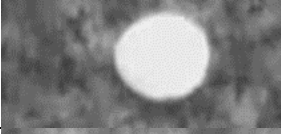


Figure 1: Sample thermal image from the UXOTi_NPA dataset.

Source: own.

Target objects can belong to one of the eleven classes (also to different classes in one image). Table 1 depicts sample UXO objects for each of the eleven classes, with some additional information.

Table 1: UXO objects from each of the eleven classes. Such objects appear in images from the evaluation dataset.

Class	Name and Type	Number of Occurrences in Dataset	RGB Image	Thermal Image
0	Mortar mine M:60 mm	139		
1	Fuse M:125 mm	161		
2/3	Hand grenade	121		
3/4	AP land mine PMR čapljinka	104		
4/5	Bullet 30.2 mm	116		
5/6	Land mine PMR 2A	136		
6/7	Mortar mine M:82 mm	149		

Class	Name and Type	Number of Occurrences in Dataset	RGB Image	Thermal Image
7/8	Land mine PMR 3	150		
8/9	Land mine PROM 1	139		
9/10	Land mine PMA 3	79		
10/2	Tromblon mine TTM RP	161		

The data splitting in the UXOT_i_NPA is prescribed. The training set contains 640 images, the validation set 80 and the testing set 88 images.

3 YOLOv7 Architecture

The You Only Look Once (YOLO) Deep Neural Network is a highly advanced object detector with an exceptional compromise between an accuracy and a speed of object detection. This is achieved by dividing the image into grid cells and considering each cell as a proposal to detect the object. Over time, many YOLO variants have been created with different features and capabilities, such as FastYOLO, YOLO-tiny, YOLO-lite, and YOLOv5n, which has been shown to be as effective as the more complex R-CNN or Faster R-CNN object detectors.

The most recent YOLO version is YOLOv7, i.e., a successor of the YOLOv5 version. YOLOv7 has typically higher accuracy than YOLOv5 but is slower when training on custom data due to its use of more floating-point operations. Additionally, it is slower on commonly available GPU systems, although it is faster on high-end ones.

We compared the effectiveness of YOLOv5 and YOLOv7 architectures on the UXOTi_NPA dataset in this study. Let us emphasize that YOLOv5 results mean the baseline results for our evaluation dataset. We experimented with five models of YOLOv5 in our previous study (Bajić & Potočnik, 2023). In this research, we compare these baseline results with the results of two YOLOv7 models, namely regular and tiny model as outlined in Table 2 (all presented metrics in Table are for the COCO dataset).

Training of the models was carried out on the Google Colaboratory platform, utilizing Tesla T4 GPUs. The programming code for the models was saved in Ultralytics' notebooks available on the Github page (Wang et al., 2022). The Pytorch framework, Python 3.7.15 and some supplementary libraries were employed to implement the models.

Table 2: Comparison of five mainstream YOLOv5 and two YOLOv7 detection models, trained and evaluated on the COCO dataset. The results were taken from (YOLOv5 Models; <https://Github.Com/Ultralytics/Yolov5>) and (Github YOLOv7).

Model	mAP@0.5	mAP@0.5:0.95	Parameters (in Million)
YOLOv5n	45.7%	28.0%	1.9
YOLOv5s	56.8%	37.4%	7.2
YOLOv5m	64.1%	45.4%	21.2
YOLOv5l	67.3%	49.0%	46.5
YOLOv5x	68.9%	50.7%	86.7
YOLOv7 tiny	56.7%	N/A	6.2
YOLOv7	69.7%	N/A	36.9

4 Results

This section presents the findings of our experiments. Initially, we solved the problem of classification and detection of UXOs from 11 different classes (i.e., an eleven UXO class detection problem). The five different models of the YOLOv5 version and two models of YOLOv7 version were trained separately on the UXOTi_NPA database's training set, for 300 epochs, utilizing identical hyperparameters. Subsequently, the trained CNNs were evaluated on the UXOTi_NPA testing set, with the classification and detection effectiveness measured using classic metrics, including Precision, Recall, Mean Average Precision (mAP) at a 0.5 threshold, and the average mAP within the 0.5 to 0.95 threshold range at an interval of 0.05. Table 4 presents results of our initial experiment, and,

at the same time, it also includes the number of free parameters (in million) for each model.

Table 3: Effectiveness of five models of the YOLOv5 and two models of YOLOv7, retrained on the UXOTi_NPA dataset: an eleven UXO class detection problem. The column 'Parameters' presents the number of CNN-free parameters.

Model	mAP@0.5	mAP@0.5:0.95	Parameters (in Million)
YOLOv5n	99.5%	87.0%	1.9
YOLOv5s	99.5%	88.5%	7.2
YOLOv5m	99.5%	89.9%	21.2
YOLOv5l	99.5%	90.5%	46.5
YOLOv5x	99.5%	89.7%	86.7
YOLOv7 tiny	99.5%	86.8%	6.2
YOLOv7	99.5%	89.7%	36.9

In the subsequent experiment, we combined UXOs from all 11 classes into a single class (i.e., 'a single UXO class detection problem'). The objective was to detect if any UXOs were present in a thermal image, irrespective of their type, and to identify their location. This scenario simulated a real-world situation where the primary interest of UXO removers is to ascertain whether UXOs are present in a given area before determining their type. The same YOLOv5 and YOLOv7 models as used in the first experiment were employed also in this experiment, only the output layer was modified accordingly. The training process and other parameters remained identical to those used in the previous experiment. The models were evaluated on the testing dataset of the UXOTi_NPA database, where the UXOs from all 11 classes were combined into a single testing class. The results are shown in Table 4.

Table 4: Effectiveness of five models of the YOLOv5 and two models of YOLOv7, retrained on the UXOTi_NPA dataset: a single UXO class detection problem. The column 'Parameters' presents the number of CNN-free parameters.

Model	mAP@0.5	mAP@0.5:0.95	Parameters (in Million)
YOLOv5n	99.5%	87.9%	1.8
YOLOv5s	99.5%	90.5%	7.1
YOLOv5m	99.5%	90.6%	21.1
YOLOv5l	99.5%	91.1%	46.4
YOLOv5x	99.5%	91.5%	86.6
YOLOv7 tiny	99.5%	86.8%	6.1
YOLOv7	99.5%	90.3%	36.8

Figure 2 and Figure 3 present some qualitative results. Figure 2 depicts results for the eleven UXO class detection problem, while results for a single UXO class detection problem are shown in Figure 3. The left column of both figures displays

the annotated original thermal images from the UXOTi_NPA testing set. Annotations (and results) are presented in a form of bounding boxes, whereat the (detected) class is specified by an index and colour (see also Table 1 for indexes). In the case of binary detection (Figure 3), all UXOs are grouped into a single class (index 1).

On the other hand, the middle and the right columns of both figures depict the UXO detection results by using our retrained YOLOv5 and YOLOv7 models.

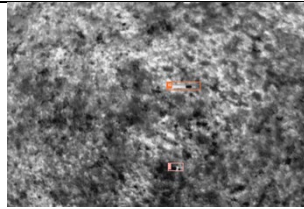
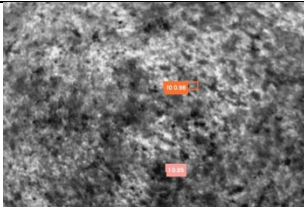
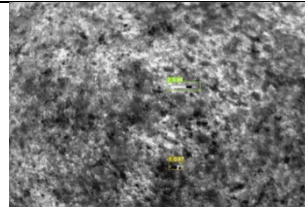
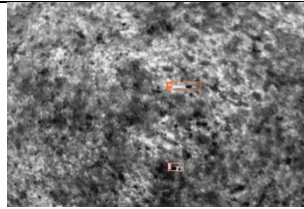
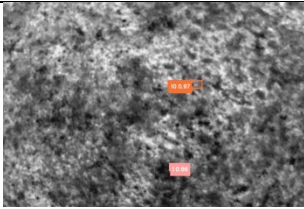
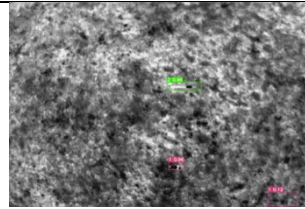
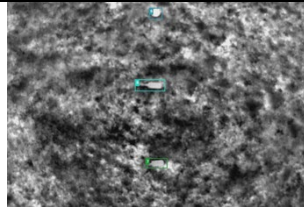
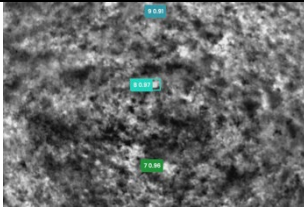
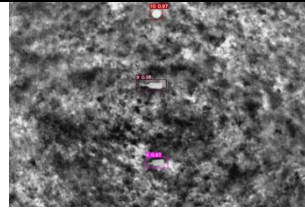
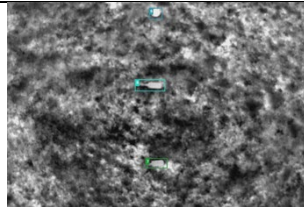
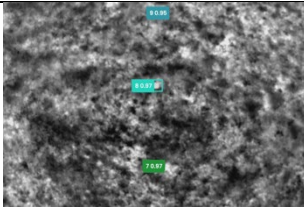
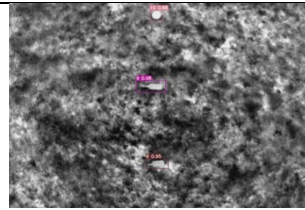
Annotated original	Detection by YOLOv5	Detection by YOLOv7
		
		
		
		

Figure 2: Qualitative results for the eleven UXO class detection problem: Expert annotations are overlaid on the thermal images (left column) and bounding boxes with UXOs, detected by using the retrained YOLOv5n and YOLOv7 tiny (rows 1 and 3, middle and right column), or YOLOv5x and YOLOv7 regular (rows 2 and 4, middle and right column). The class index and detection probability are written next to the bounding box.

Source: own.

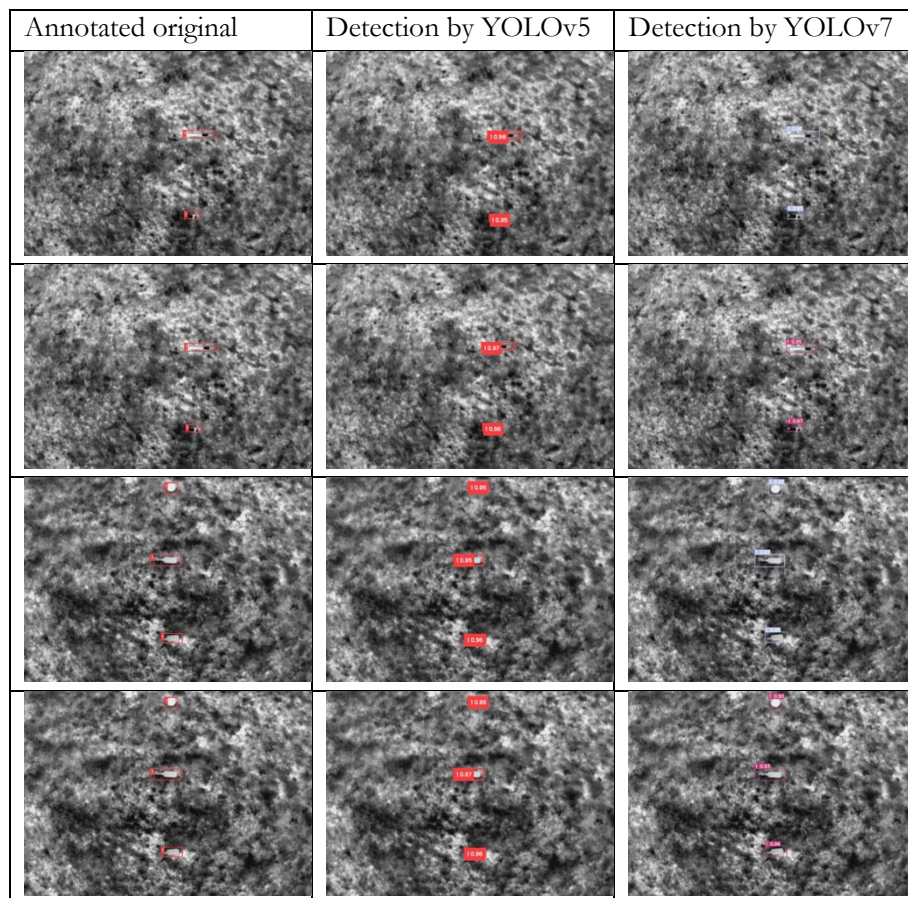


Figure 3: Qualitative results for the single UXO class detection problem: Expert annotations are overlaid on the thermal images (left column) and bounding boxes with UXOs, detected by using the pretrained YOLOv5n and YOLOv7 tiny (rows 1 and 3, middle and right column) or YOLOv5x and YOLOv7 regular (rows 2 and 4, middle and right column). A single class index 1 and detection probability are written next to the bounding box.

Source: own.

5 Discussion and Conclusion

In this study, we have experimented with a state-of-the-art computational approach utilizing Deep Learning and Convolutional Neural Networks aimed to detect Unexploded Ordnances (UXOs) from thermal imaging materials. The detection of UXOs is an essential support activity by the neutralization of explosive war remnants. With our automated solution, a terrain can be examined from the air (e.g., by using Unmanned Aerial Vehicles) and potentially dangerous areas can be identified in advance.

Our study's innovation is the utilization of thermal images for UXO detection. Thermal imaging provides valuable information about the environment and its changes, particularly if this information is not detectable in the visible spectrum. Different materials, including UXOs, leave their unique thermal signatures in the image, making it possible to identify them with high accuracy. An example is a land mine with a green metal casing placed on green grass. Such a mine is impossible to detect in the visible spectrum, but it is easily separable in the thermal spectrum (see Figure 4).

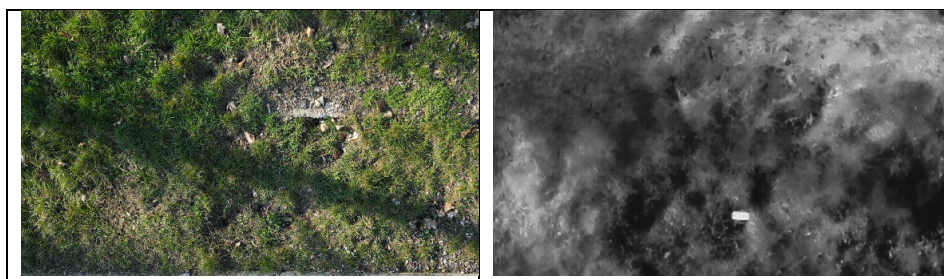


Figure 4. A land mine with a green metal casing placed on green grass: An image in the visible spectrum (left) and the same terrain in the thermal spectrum (right). Note: This example is from our private database and is not part of the UXOTi_NPA dataset.

Source: own.

We also compared the performance of YOLOv7 with that of YOLOv5 on the UXOTi_NPA public dataset in this article. YOLOv5 is an established architecture thoroughly tested on various datasets, while YOLOv7 was introduced recently (in 2022) and is still in the development phase. Based on the obtained standardised metrics mAP on the UXOTi_NPA dataset, we conclude that the YOLOv5 seems

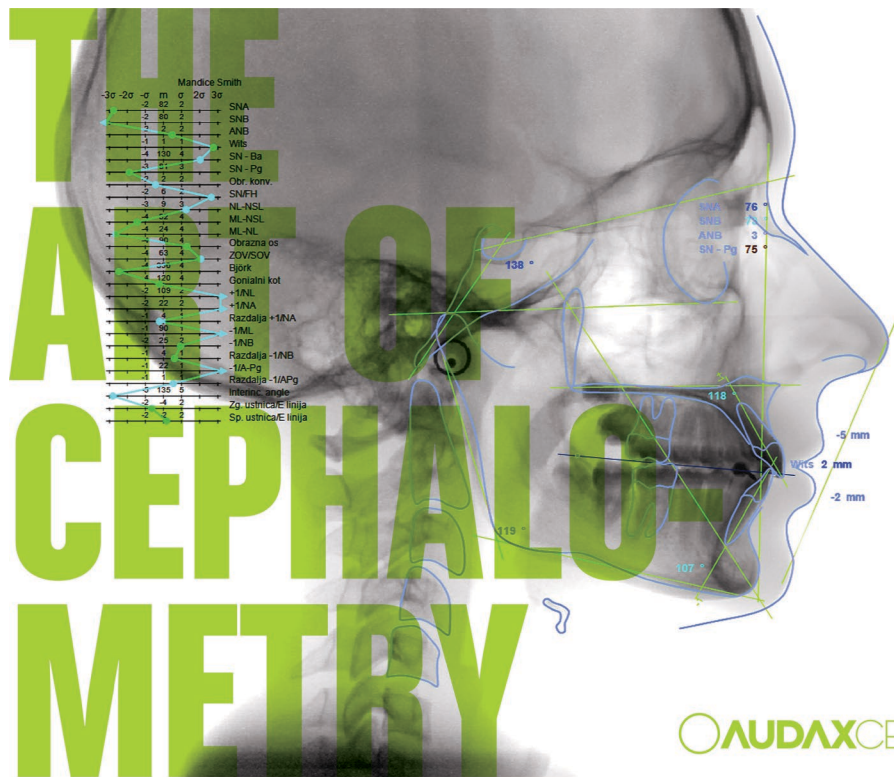
to be a better trade-off between model size and detection effectiveness than YOLOv7. The reason lies undoubtedly in the higher number of false positive detection by YOLOv7 (see Figure 2). We also see a possible cause in a small UXOTi_NPA dataset. Considering a training time on the UXOTi_NPA dataset, YOLOv7 requires more processing time and resources compared to YOLOv5, whereat such an increase was not justified by the results. A similar discussion can be found in (Durve et al., 2023).

In this work, we have shown that the combination of deep Convolutional Neural Networks and thermal imaging can be used advantageously to detect UXOs in a real environment. Our future work will go in several directions, primarily in an expansion of the UXOTi_NPA dataset with new objects and environments, and in an upgrade of YOLOv7 with a module for the more efficient feature analysis and fusion.

References

- Bajic, M. (2020). Testing of remotely piloted aircraft systems with a thermal infrared camera to detect explosive devices at contaminated areas and validation of developed standard operational procedures. Norwegian Peoples Aid Oslo Norway.
- Bajić, M., & Potočnik, B. (2023). UAV Thermal Imaging for Unexploded Ordnance Detection by Using Deep Learning. *Remote Sensing*, 15(4), 967. <https://doi.org/10.3390/rs15040967>
- Banuls, A., Mandow, A., Vazquez-Martin, R., Morales, J., & Garcia-Cerezo, A. (2020). Object Detection from Thermal Infrared and Visible Light Cameras in Search and Rescue Scenes. 2020 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR), 380–386. <https://doi.org/10.1109/SSRR50563.2020.9292593>
- Baur, J., Steinberg, G., Nikulin, A., Chiu, K., & de Smet, T. S. (2020). Applying Deep Learning to Automate UAV-Based Detection of Scatterable Landmines. *Remote Sensing*, 12(5), 859. <https://doi.org/10.3390/rs12050859>
- Dai, X., Yuan, X., & Wei, X. (2021). TIRNet: Object detection in thermal infrared images for autonomous driving. *Applied Intelligence*, 51(3), 1244–1261. <https://doi.org/10.1007/s10489-020-01882-2>
- Durve, M., Orsini, S., Tiribocchi, A., Montessori, A., Tucny, J.-M., Lauricella, M., Camposeo, A., Pisignano, D., & Succi, S. (2023). Benchmarking YOLOv5 and YOLOv7 models with DeepSORT for droplet tracking applications. <https://doi.org/10.48550/ARXIV.2301.08189>
- GICHHD. (2022). Explosive Ordnance Guide for Ukraine—Second Edition; https://www.gichd.org/fileadmin/GICHHD-resources/rec-documents/GICHHD_Ukraine_Guide_2022_Second_Edition_web.pdf. GICHHD. https://www.gichd.org/fileadmin/GICHHD-resources/rec-documents/GICHHD_Ukraine_Guide_2022_Second_Edition_web.pdf
- Github YOLOv7. (n.d.). Retrieved February 12, 2023, from <https://github.com/WongKinYiu/yolov7>
- Krause, P., Salahat, E., & Franklin, E. (2018). Diurnal Thermal Dormant Landmine Detection Using Unmanned Aerial Vehicles. IECON 2018 - 44th Annual Conference of the IEEE Industrial Electronics Society, 2299–2304. <https://doi.org/10.1109/IECON.2018.8591378>

- Kuenzer, C., & Dech, S. (Eds.). (2013). *Thermal infrared remote sensing: Sensors, methods, applications*. Springer. <https://doi.org/10.1007/978-94-007-6639-6>
- Leira, F. S., Helgesen, H. H., Johansen, T. A., & Fossen, T. I. (2021). Object detection, recognition, and tracking from UAVs using a thermal camera. *Journal of Field Robotics*, 38(2), 242–267. <https://doi.org/10.1002/rob.21985>
- Nikulin, A., de Smet, T., Baur, J., Frazer, W., & Abramowitz, J. (2018). Detection and Identification of Remnant PFM-1 ‘Butterfly Mines’ with a UAV-Based Thermal-Imaging Protocol. *Remote Sensing*, 10(11), 1672. <https://doi.org/10.3390/rs10111672>
- Roberts, S., & Williams, J. (1995). *After the guns fall silent: The enduring legacy of landmines*. Veterans of America Foundation, Washington D.C., USA.
- Wang, C.-Y., Bochkovskiy, A., & Liao, H.-Y. M. (2022). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. <https://doi.org/10.48550/ARXIV.2207.02696>
- Yao, Y., Wen, M., & Wang, Y. (2019). Multi-Temporal IR Thermography For Mine Detection. 2019 10th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp), 1–4. <https://doi.org/10.1109/Multi-Temp.2019.8866906>
- YOLOv5 models; <https://github.com/ultralytics/yolov5>. (n.d.). Retrieved September 1, 2022, from <https://github.com/ultralytics/yolov5>



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

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

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

AUDAXCEPH®

www.audaxceph.com

info@audaxceph.com

AUDAX d.o.o. • Bežovškova pot. 10 • SI-1000 Ljubljana
Tel: +386 1 200 40 50 • Fax: +386 1 423 41 00

**SPS V
ŠTEVILKAH****Σ REZULTATI
2022**

"V letu 2022 je
SPS pomagal
4.605
podjetjem z
154,64
mio EUR
finančnimi in
vsebinskimi
pomočmi"

**Σ PLAN
2023**

"V letu 2023
SPS načrtuje
pomagati
2.500
podjetjem z
140
mio EUR
finančnimi in
vsebinskimi
pomočmi"

»SPS je vodilni finančni sklad za podporo MSP-jev, vključno s start-upi in scale-upi, ki s:

FINANČNIMI
SPODBUDAMISPODBUDAMI ZA DVIG
KOMPETENCSPODBUDAMI ZA
TRAJNOSTNI RAZVOJ

zmanjšuje vrzeli, ki so jih podjetja deležna na trgu, hkrati pa pri svojem poslovanju upošteva načelo ESG.«

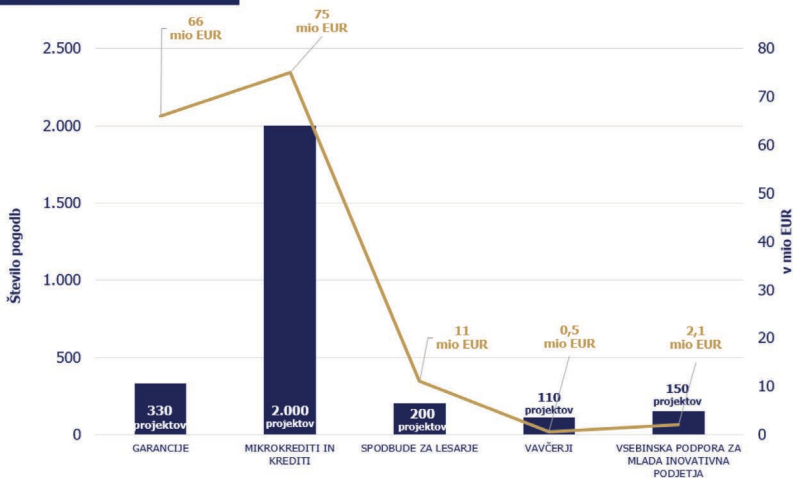
I. VLOGA SPSa V LETU 2023**V okolju SPS igra pomembno vlogo, ker nastopa kot:**

1. Ponudnik garancij za bančne kredite
2. Ponudnik mikrokreditov in specifičnih kreditnih linij
3. Sooblikovalec start-up in scale-up ekosistema v Sloveniji
4. Ključni soinvestitor semenskega in tveganega kapitala v Sloveniji
5. Ponudnik osnovnih spodbud za dvig kompetenc (skills) v MSP-jih
6. Partner pri spodbujanju trajnostnega razvoja v Sloveniji
7. Partner pri spodbujanju trajnostnega razvoja na mednarodnem nivoju
8. Učinkovita javna finančna institucija

Plan finančne pomoči v letu 2022:

- o **140 mio EUR** novih pomoči
- o **cca 2.500 podjetij**, ki bodo v letu 2023 koristila pomoč SPSa
- o pri tem SPS ocenjuje, da bodo vložena sredstva spodbudila cca 237,00 mio EUR novih investicij v gospodarstvu, kar je primerljivo s preteklimi leti oz. z leti pred pandemijo COVID

Pregled spodbud 2023:



Pregled po produktih 2023:

PRODUKTI 2022	NAČRTOVANA VIŠINA SREDSTEV	PREDVIDENA OBJAVA
1. Garancije – P1 plus 2023	66 mio EUR	Marec 2023
2. Mikrokrediti – P7 2023 I	20 mio EUR	Februar/ Marec 2023
3. Mikrokrediti – P7 2023 II	10 mio EUR	Junij/Julij 2023
4. Krizni mikrokrediti – P7-2 KRIZNI 2023	5 mio EUR	Februar/ Marec 2023
5. Krizno likvidnostni krediti – P7CE 2023 I	10 mio EUR	Marec 2023
6. Krizno likvidnostni krediti – P7CE 2023 II	30 mio EUR	Marec 2023
7. Spodbude za lesarstvo – P4L 2023	11 mio EUR	Marec 2023
8. Vavčerji	0,5 mio EUR	v teku
9. Vsebinska podpora za mlada inovativna podjetja	2,10 mio EUR	v teku

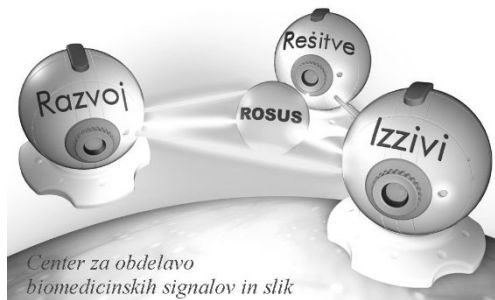
POKROVITELJI

Audax

**SLOVENSKI
PODJETNIŠKI
SKLAD**



Slovenska sekcija



ROSUS 2023

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

ROSUS 2023 - RAČUNALNIŠKA OBDELAVA SLIK IN NJENA UPORABA V SLOVENIJI 2023: ZBORNİK 17. STROKOVNE KONFERENCE

BOŽIDAR POTOČNIK (UR.)

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

Povzetek ROSUS 2023 – Računalniška obdelava slik in njena uporaba v Sloveniji 2023 je strokovna računalniška konferenca, ki jo od leta 2006 naprej vsako leto organizira Inštitut za računalništvo iz Fakultete za elektrotehniko, računalništvo in informatiko, Univerze v Mariboru. Konferenca povezuje strokovnjake in raziskovalce s področij digitalne obdelave slik in strojnega vida z uporabniki tega znanja, pri čemer uporabniki prihajajo iz raznovrstnih industrijskih okolij, biomedicine, športa, zabavnštva in sorodnih področij. Zbornik konference ROSUS 2023 združuje strokovne prispevke več avtorjev, od tega štiri vabljena predavanja 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.

Ključne besede:

računalniška
obdelava
slik
strojni
vid,
biomedicina,
industrijske
aplikacije,
prenos
znanja

ROSUS 2023 - COMPUTER IMAGE PROCESSING AND ITS APPLICATION IN SLOVENIA 2023: PROCEEDINGS OF THE 17TH PROFESSIONAL CONFERENCE

BOŽIDAR POTOČNIK (ED.)

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

Abstract ROSUS 2023–Computer image processing and its application in Slovenia 2023 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 2023 combine scientific articles by dozens of authors, including four 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.

Keywords:

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



Univerza v Mariboru

Fakulteta za elektrotehniko,
računalništvo in informatiko

LSPO

Laboratorij za sistemsko
programsko opremo

<https://rosus.feri.um.si/rosus2023>
rosus.feri@um.si

