



LATVIJAS
UNIVERSITĀTE

Promocijas darba
kopsavilkums

Maksims Ivanovs

DZIĻĀS MAŠĪNMĀCĪŠANĀS
PIELIETOJUMI LIETIŠKAJĀ
DATORREDZĒ:
ATTĒLU IZPRATNES
UZDEVUMU RISINĀŠANA AR
KONVOLŪCIJU NEIRONU
TĪKLIEM

Rīga 2025



LATVIJAS
UNIVERSITĀTE

**EKSAKTO ZINĀTŅU UN
TEHNOLOĢIJU FAKULTĀTE**

Maksims Ivanovs

**DZIĻĀS MAŠĪNMĀCĪŠANĀS PIELIETOJUMI
LIETIŠĶAJĀ DATORREDZĒ:
ATTĒLU IZPRATNES UZDEVUMU RISINĀŠANA
AR KONVOLŪCIJU NEIRONU TĪKLIEM**

PROMOCIJAS DARBA KOPSAVILKUMS

Doktora grāda (*Ph. D.*) iegūšanai
datorzinātnes nozarē
Apakšnozare: mākslīgais intelekts

Rīga 2025

Promocijas darbs izstrādāts Elektronikas un datorzinātņu institūtā (EDI) un Latvijas Universitātes Eksakto zinātņu un tehnoloģiju fakultātes Datorikas nodaļā laika posmā no 2019. līdz 2024. gadam.

NACIONĀLAIS
ATTĪSTĪBAS
PLĀNS 2020



EIROPAS SAVIENĪBA
Eiropas Sociālais
fonds

SAM 8.2.2. 'Akadēmiskā personāla atjaunotne un kompetenču pilnveide Latvijas Universitātē', projekta Nr. 8.2.2.0/18/A/010.

Darbs sastāv no ievada, piecām nodaļām, secinājumiem un literatūras saraksta.

Darba forma: disertācija datorzinātnes nozarē, mākslīgā intelekta apakšnozarē.

Darba zinātniskais vadītājs: vadošais pētnieks, Dr. sc. ing. **Roberts Kadiķis**.

Darba recenzenti:

1. Profesors, Dr. sc. comp. **Jānis Zuters**, Latvijas Universitāte.
2. Asociētais profesors, Dr. sc. ing. **Andrejs Romānovs**, Rīgas Tehniskā universitāte.
3. Lektors, PhD **Mahmuds Elbatahs** (Mahmoud Elbattah), Mākslas, tehnoloģiju un vides koledža, Rietumanglijas Universitāte, Bristole, Lielbritānija un MIS laboratorija, Pikardijas Žila Verna universitāte, Amjēna, Francija.

Promocijas darba aizstāvēšana notiks 2025. gada 16. maijā Latvijas Universitātes Datorzinātnes un informātikas promocijas padomes atklātā sēdē.

Ar promocijas darbu un tā kopsavilkumu var iepazīties Latvijas Universitātes Bibliotēkā Rīgā, Kalpaka bulvārī 4.

Promocijas padomes priekšsēdētājs _____ / **Guntis Bārzdīņš** /

Promocijas padomes sekretāre _____ / **Sintija Siliņa** /

© Latvijas Universitāte, 2025

© Maksims Ivanovs, 2025

ISBN 978-9934-36-372-6

ISBN 978-9934-36-373-3 (PDF)

Anotācija

Promocijas darbs aplūko dziļās mašīnmācīšanās metožu pielietojumus svarīgu attēlu izpratnes uzdevumu – attēlu klasifikācijas, objektu atpazīšanas un semantiskās segmentācijas – risināšanai.

Literatūras apskatā pirmajā nodaļā es sniedzu pamatojumu pārējās disertācijas nodaļās izklāstītajam eksperimentālajam darbam, noskaidroju, ka visperspektīvākais dziļo neironu tīklu arhitektūras veids attēlu izpratnes uzdevumiem ir konvolūciju neironu tīkli (KNT), kā arī apspriežu ar datu pieejamību KNT apmācībai saistītas problēmas un sniedzu ieskatu tajā, kā zināšanu pārnese un sintētiskie dati var tikt izmantoti, lai mazinātu šīs problēmas.

Darba praktiskā daļa aplūko KNT pielietojumus lietišķo attēlu izpratnes problēmu risināšanai. Otrajā nodaļā es apskatu pētījumus par KNT izmantošanu roku mazgāšanas kustību atpazīšanai, lai izstrādātu sistēmu roku higiēnas novērošanai. Trešajā nodaļā es aprakstu KNT izmantošanu ielu skatu semantiskai segmentācijai, kas ir svarīgs uzdevums pašbraucošo automašīnu navigācijas sistēmām. Ceturtajā nodaļā es aprakstu KNT balstītu objektu detektoru izmantošanu plastmasas pudeļu atpazīšanai, lai tās varētu paņemt robota roka. Visbeidzot, piektajā nodaļā es apskatu KNT pielietojumu mikroskopijas attēlu klasifikācijai, lai automatizētu orgānu uz čipa audzēšanas novērošanu.

Darba mērķis – piedāvāt efektīvus risinājumus lietišķiem attēlu izpratnes uzdevumiem – tika sasniegts visos uzdevumos, izņemot klasifikācijas eksperimentus uz PSKUS datu kopas, kas ir vissarežģītākā un trokšņainākā roku mazgāšanas video datu kopa. Turklāt darbā izklāstītie rezultāti veicināja labāku izpratni par metodoloģiskajiem izaicinājumiem dziļās mašīnmācīšanās jomā, piemēram, par pieejām reālās pasaules datu kopu papildināšanai ar sintētiskiem datiem. Darbā aprakstītie rezultāti ir publicēti sešos zinātniskos rakstos, kas indeksēti *Elsevier Scopus* un/vai *Web of Science* datubāzēs, un prezentēti četrās konferencēs. Rezultātu aprobācija tika veikta septiņos zinātniskajos projektos, kas īstenoti Elektronikas un datorzinātņu institūtā (EDI), kurā tika izstrādāts šis darbs. Darba rezultāti pamato četras aizstāvēšanai izvīzītās tēzes.

Saturs

Ievads	5
1. Teorētiskais apskats	9
1.1. Attēlu izpratnes uzdevumu risināšanas metodes	9
1.2. Datu kopas attēlu izpratnes uzdevumiem	10
2. Roku mazgāšanas kustību klasi ikācija	12
2.1. Datu kopas	12
2.1.1. PSKUS datu kopa	12
2.1.2. METC datu kopa	13
2.2. Sākotnējie eksperimenti uz PSKUS un METC datu kopām .	13
2.3. KNT efektivitātes izpēte uz dažādām datu kopām	14
2.3.1. Metodoloģija	14
2.3.2. Rezultāti	15
3. Ielu skatu semantiskā segmentācija	17
3.1. Datu kopas	17
3.2. Metodoloģija	18
3.3. Rezultāti	18
4. Objektu atpazīšana groza izkraušanas uzdevuma veikšanai	22
4.1. Datu kopas	22
4.2. Metodoloģija	23
4.3. Rezultāti	23
5. Attēlu klasifikācija orgānu uz čipa audzēšanas novērošanai	25
5.1. Eksperimenti uz sākotnējās orgānu uz čipa attēlu datu kopas	25
5.1.1. Dati un metodoloģija	25
5.1.2. Rezultāti	26
5.2. Eksperimenti uz orgānu uz čipa gala datu kopas	26
5.2.1. Dati un metodoloģija	26
5.2.2. Rezultāti	27
Secinājumi	29
Literatūras saraksts	31
Pateicības	35

Ievads

Datorredze, kas sāka attīstīties 20. gadsimta 60. gadu beigās, mūsdienās ir plaša un viena no aktīvākajām pētniecības nozarēm datorzinātnē, jo automatizētām sistēmām, kas spēj saprast vizuālas ainas, ir daudz iespējamu pielietojumu. Šajā darbā es pievērsos trim attēlu izpratnes problēmu veidiem: attēlu klasifikācijai, attēlu segmentācijai un objektu atpazīšanai. Mans darbs pie šo problēmu risināšanas pieder lietišķās datorredzes jomai, tas ir, datorredzes metožu pielietošanas praktiskiem uzdevumiem zinātnē un nozarē. Es izmantoju datorredzes metodes, lai risinātu šādus reālās pasaules attēlu izpratnes uzdevumus:

- roku mazgāšanas kustību klasifikācija klīniskos apstākļos, lai novērotu roku higiēnu (2. nodaļa);
- ielu skatu semantiskā segmentācija, lai uzlabotu pašbraucošo auto navigācijas sistēmas (3. nodaļa);
- satveramo pudeļu atpazīšana pudeļu kaudzē, lai robota roka varētu veikt groza izkraušanas uzdevumu (4. nodaļa);
- mikroskopijas attēlu klasifikācija, lai automatizētu orgānu uz čipa audzēšanas novērošanu (5. nodaļa).

Visos šajā darbā aplūkotajos pētījumos es risinu attēlu izpratnes uzdevumus ar dziļās mašīnmācīšanās palīdzību, t.i., izmantojot dziļos neironu tīklus (DNT), kuri pašlaik tiek uzskatīti par efektīvāko risinājumu vairumam datorredzes uzdevumu. Ņemot vērā, ka DNT tiek veiksmīgi pielietoti arī daudzās citās mākslīgā intelekta (MI) jomās, vispārīgu dziļās mašīnmācīšanās metodoloģisko izaicinājumu risināšana šajā darbā – piemēram, nepieciešamība pēc lielām datu kopām modeļu apmācībai, kvalitātes kontroles problēmas lielapjoma datu kopās un sintētisko datu izmantošana reālās pasaules datu kopu paplašināšanai – var veicināt arī šo jomu attīstību.

Darba mērķis ir piedāvāt efektīvus risinājumus lietišķiem attēlu izpratnes uzdevumiem. **Darba galvenais pieņēmums** ir, ka konvolūciju neironu tīkli (KNT) – DNT veids, kas īpaši piemērots perceptuālo datu apstrādei – var veiksmīgi atrisināt šajā darbā apskatītos attēlu izpratnes uzdevumus. **Darba uzdevumi** un **hipotēzes** ir atkarīgi no konkrētā uzdevuma un tādēļ ir definēti šī darba attiecīgās nodaļās. **Pētījuma metodes**, ko es izmantoju šajā darbā, ir tās, kuras parasti tiek pielietotas MI un datorredzes pētījumos: literatūras izpēte un analīze, datu tīrīšana un pirmapstrāde, sintētisko datu ģenerēšana, eksperimentu ar DNT plānošana un veikšana, kā arī eksperimentu rezultātu analīze un validācija.

Šajā disertācijā veikto pētījumu rezultātā, es izvirzu aizstāvēšanai sekojošas tēzes:

- **Pirmā tēze:** Pielietojot KNT reālās pasaules attēlu izpratnes uzdevumiem, datu pieejamība un kvalitāte rada lielākus izaicinājumus nekā modeļa izvēle un pielāgošana.
- **Otrā tēze:** Konvolūciju neironu tīkliem, kas uzrāda labus rezultātus, kad tie apmācīti un izvērtēti uz laboratorijas apstākļos iegūtām datu kopām, var būt grūti sasniegt līdzīgus panākumus, kad tie apmācīti un izvērtēti uz sarežģītākiem reālās pasaules datiem.
- **Trešā tēze:** Lai gan modernākie KNT balstītie attēlu klasifikatori un objektu detektori ar lielāku parametru skaitu parasti demonstrē augstāku precizitāti uz salīdzinošai novērtēšanai (angl. *benchmarking*) paredzētām datu kopām nekā klasifikatori un objektu detektori ar mazāku parametru skaitu, šī starpība samazinās vai pat pazūd, kad modeļi tiek apmācīti un novērtēti uz mazākām un sarežģītākām reālās pasaules datu kopām.
- **Ceturrtā tēze:** Lai gan reālās pasaules datu kopu papildināšana ar fotoreālistiskiem sintētiskiem attēliem ir efektīvs veids, kā uzlabot uz šādiem datiem apmācītu KNT precizitāti, sintētisko datu apjoma palielināšana tieši nekorelē ar precizitātes uzlabojumiem attēlu izpratnes uzdevumos.

Šīs tēzes galvenokārt balstītas uz rezultātiem šādās darba nodaļās: pirmā tēze – 2., 3., 4. un 5. nodaļā, otrā tēze – 2. nodaļā; trešā tēze – 4. un 5. nodaļā; ceturrtā tēze – 3. un 5. nodaļā.

Promocijas darbā aprakstītie pētījumu rezultāti ir publicēti šādos zinātniskos rakstos, kas indeksēti *Elsevier Scopus* un/vai *Web of Science* datubāzēs:

1. M. Lulla, A. Rutkovskis, A. Slavinska, A. Vilde, A. Gromova, **M. Ivanovs**, A. Skadins, R. Kadikis, A. Elsts, “Hand-washing video dataset annotated according to the World Health Organization’s hand-washing guidelines,” *Data*, vol. 6, no. 4:38, 2021.
2. A. Elsts, **M. Ivanovs**, R. Kadikis, O. Sabelnikovs, “CNN for hand washing movement classification: What matters more – the approach or the dataset?,” in *2022 Eleventh International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pp. 1–6, IEEE, 2022.
3. **M. Ivanovs**, K. Ozols, A. Dobrajs, R. Kadikis, “Improving semantic segmentation of urban scenes for self-driving cars with synthetic images,” *Sensors*, vol. 22, no. 6:2252, 2022.
4. D. Duplevska, **M. Ivanovs**, J. Arents, R. Kadikis, “Sim2Real image translation to improve a synthetic dataset for a bin picking task,” in

2022 IEEE 27th International Conference on Emerging Technologies and Factory Automation (ETFA), pp. 1–7, IEEE, 2022.

5. **M. Ivanovs**, L. Leja, K. Zviedris, R. Rimša, K. Narbutė, V. Movčana, F. Rumnieks, A. Strods, K. Gillois, G. Mozolevskis, A. Abols, R. Kadikis, “Synthetic image generation with a fine-tuned latent diffusion model for organ on chip cell image classification,” in *2023 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)*, pp. 1–6, IEEE, 2023.
6. V. Movčana, A. Strods, K. Narbutė, F. Rūmnieks, R. Rimša, G. Mozolevskis, **M. Ivanovs**, R. Kadikis, K. Zviedris, L. Leja, A. Zujeva, T. Laimiņa, A. Abols, “Organ-On-A-Chip (OOC) Image Dataset for Machine Learning and Tissue Model Evaluation,” *Data*, vol. 9, no. 2:28, 2024.

Promocijas darbā aprakstītie pētījumu rezultāti ir publicēti šādos zinātniskos rakstos, kas nav indeksēti *Elsevier Scopus* un/vai *Web of Science* datubāzēs:

1. **M. Ivanovs**, R. Kadikis, M. Lulla, A. Rutkovskis, and A. Elsts, “Automated quality assessment of hand washing using deep learning,” *arXiv:2011.11383*, 2020.
2. O. Zemlanuhina, M. Lulla, A. Rutkovskis, A. Slavinska, A. Vilde, A. Melbarde-Kelme, A. Elsts, **M. Ivanov**, and O. Sabelnikovs, “Influence of different types of real-time feedback on hand washing quality assessed with neural networks/simulated neural networks,” in *SHS Web of Conferences*, vol. 131, p. 02008, EDP Sciences, 2022.

Esmu prezentējis šajā disertācijā aprakstītos pētījumu rezultātus sekojošās konferencēs:

1. IEEE International Conference on Image Processing Theory, Tools & Applications – IPTA 2022, Zalcburga, Austrija, 2022.
Prezentācija *CNN for Hand Washing Movement Classification: What Matters More — the Approach or the Dataset?*
2. 27th IEEE International Conference on Emerging Technologies and Factory Automation – EFTA 2022, Štutgarte, Vācija, 2022.
Prezentācija *Sim2Real Image Translation to Improve a Synthetic Dataset for a Bin Picking Task*.
3. SEMICON Europa 2022 Conference, Minhene, Vācija, 2022.
Prezentācija *Synthetic Data for Robotics: Opportunities and Challenges*.
4. 26th IEEE Signal Processing: Algorithms, Architectures, Arrangements and Applications Conference – SPA 2023, Poznaņa, Polija, 2023.

Prezentācija *Synthetic Image Generation With a Fine-Tuned Latent Diffusion Model for Organ on Chip Cell Image Classification*.

Promocijas darba ietvaros veiktie pētījumi tika īstenoti un aprobēti Elektronikas un datorzinātņu institūtā (EDI). Šie pētījumi bija daļa no vairākiem zinātniskiem projektiem EDI un tika atbalstīti ar šo projektu finansējumu. Projektu saraksts ir sekojošs:

1. *Programmable Systems for Intelligence in Automobiles – PRYSTINE* (Horizon 2020 ECSEL Joint Undertaking funding under grant agreement 783190).
2. *Efficient module for automatic detection of people and vehicles using video surveillance cameras – VAPI* (ERDF project No. 1.2.1.1/18/A/006 research No. 1.5).
3. *Automated hand washing quality control and hand washing quality evaluation system with real-time feedback – Handwash* (project No. lzp-2020/2-0309).
4. *Integration of reliable technologies for protection against Covid-19 in healthcare and high risk areas – COV-CLEAN* (project No. VPP-COVID-2020/1-004).
5. *Intelligent Motion Control under Industry 4.E – IMOCO4.E* (Horizon 2020 ECSEL Joint Undertaking funding under grant agreement 101007311).
6. *AI-Improved Organ on Chip Cultivation for Personalised Medicine – AImOOC*
(contract with Central Finance and Contracting Agency of Republic of Latvia no. 1.1.1.1/21/A/079; the project was cofinanced by REACT-EU funding for mitigating the consequences of the pandemic crisis).
7. *Holographic microscopy- and artificial intelligence-based digital pathology for the next generation of cytology in veterinary medicine – VetCyto* (project No. lzp-2023/1-0220).

Darbs sastāv no ievada, piecām nodaļām, secinājumiem un izmantotās literatūras saraksta. Darba garums ir 146 lappuses.

1. Teorētiskais apskats

Šajā darbā es izmantoju [1] sniegto **datorredzes** definīciju kā “paņēmienu kopumu, lai iegūtu, apstrādātu, analizētu un izprastu sarežģītus daudzdimensiju [vizuālos] datus no apkārtējās vides”. Sekojot [2] pieejai, es nošķiru datorredzi no digitālo attēlu apstrādes, jo digitālā attēlu apstrāde attiecas uz zema līmeņa darbībām ar attēliem, piemēram, attēlu uzlabošanu vai krāsu apstrādi, savukārt plašāks jēdziens “datorredze” ietver arī augsta līmeņa darbības ar vizuāliem datiem.

Ņemot vērā to, ka datorredze aptver vairākas disciplīnas un plašu pētniecības problēmu loku, es ieviešu šaurāku terminu, lai apzīmētu pētniecības problēmas, ar kurām galvenokārt ir saistīts mans darbs. Šim nolūkam es izmantoju terminu **attēlu izpratne**, kā to definē [3], proti, kā metožu un tehniku kopums, kas “cenšas interpretēt attēla nozīmi augstā līmenī, lai sniegtu semantisku informāciju, kas būtu tuva cilvēka domāšanai un ļautu pieņemt lēmumus un vadīt darbības atbilstoši ainu izpratnei”. Mani pētījumi, kas veido šī darba pamatu, ir saistīti ar trim nozīmīgiem attēlu izpratnes uzdevumiem: **attēlu klasifikāciju**, **objektu atpazīšanu** un **attēlu segmentāciju**.

Attēlu klasifikācija būtībā ir marķēšana [4]: klasifikators nosaka attēla I piederību klasei C_i , kas ir elements fiksētā izskatāmo klašu kopā $C = \{C_1, C_2, C_3, \dots, C_N\}$.

Objektu atpazīšana ir sarežģītāks uzdevums nekā attēlu klasifikācija, jo objektu detektors gan piešķir klasi $C_i \in C$ katram objektam attēlā, gan nosaka objektu atrašanās vietas. Citiem vārdiem sakot, kamēr attēlu klasifikatora mērķis ir atbildēt uz jautājumu *Kāds objekts tur ir?*, objektu detektors atbild uz jautājumu *Kur un kādi objekti tur ir?* [5].

Attēlu segmentācijas mērķis ir sadalīt ievadattēlu vairākos segmentos (t.i., nepārtrauktās pikseļu grupās) jēgpilnā veidā. Viena no galveniem attēlu segmentācijas veidiem – **semantiskās segmentācijas** – mērķis ir piešķirt katram iegūto segmentu pikselim vienu klasi $C_i \in C$. Tādējādi semantiskā segmentācija būtībā ir daudzklašu klasifikācija pikseļu līmenī; ņemot vērā to, ka tā tiek veikta katram ievadattēla pikselim, tā parasti tiek uzskatīta par sarežģītāku uzdevumu nekā attēlu klasifikācija [6].

1.1. Attēlu izpratnes uzdevumu risināšanas metodes

Attēlu izpratnes uzdevumu risināšanas metodes kopumā var iedalīt divās galvenajās kategorijās: klasiskās metodes un dziļajā mašīnmācīšanā balstītās metodes.

Klasisko attēlu izpratnes metožu noteicoša iezīme ir tas, ka tās pilnībā vai daļēji balstās uz tieši programmētiem algoritmiem un paļaujas uz jomas ekspertu atlasītām datu iezīmēm un noteikumiem. Lai gan šīs metodes joprojām dažreiz tiek izmantotas mūsdienās, klasisko pieeju vietā parasti tiek izvēlēti DNT [7], jo tie ir precīzāki un robustāki.

Dziļajā mašīnmācīšanās balstīto metožu popularitātes pieaugums sākās 2012. gadā, kad KNT AlexNet [8] uzvarēja ILSVRC attēlu klasifikācijas sacensībās [9]. Dažu gadu laikā dziļajā mašīnmācīšanās balstītās metodes ieguva dominējošo pozīciju attēlu klasifikācijas jomā, un līdzīgs pavērsiens notika arī objektu atpazīšanas un semantiskās segmentācijas jomās. Tādējādi dziļā mašīnmācīšanās pašlaik ir ļoti plašs un strauji augošs pētniecības lauks gan zinātnē, gan nozarē ar plašu pieejamo modeļu arhitektūru klāstu. **KNT**, kas ir populārākā arhitektūra attēlu izpratnes uzdevumiem, ir krasi uzlabojuši mākslīgo neironu tīklu (MNT) izmantošanu datorredzē [10], kā arī ir galvenais rīks, kas tika izmantots šajā darbā aprakstītajos pētījumos. KNT risina ar telpisko attiecību attēlos izmantošanu saistītas problēmas efektīvāk, nekā to dara vienkāršāki DNT, tādi kā pilnībā savienoti vienvirziena neironu tīkli. Turklāt konvolūciju izmantošana mazina savienojumu blīvumu tīklā, tādējādi uzlabojot efektivitāti.

Pieejamo KNT arhitektūru klāsts ir ļoti plašs, sākot ar agrīno LeNet [11] līdz pat jaudīgiem jaunākiem modeļiem. Šajā darbā aprakstītos pētījumos es izmantoju MobileNetV2 [12], MobileNetV3Large [13] un EfficientNet-B7 [14] attēlu klasifikācijai, ar DeepLabv3 [15] segmentācijas bloku paplašinātos Xception [16] un MobileNetV2 semantiskai segmentācijai un YOLOv5 [17] objektu atpazīšanai.

1.2. Datu kopas attēlu izpratnes uzdevumiem

Viens no nozīmīgākajiem faktoriem dziļās mašīnmācīšanās attīstībā, it īpaši attēlu izpratnes metožu attīstībā, ir lielo datu kopu pieejamība modeļu apmācībai, novērtēšanai un salīdzinošai vērtēšanai [18, 19]. Brīvpiekļuve šādām datu kopām ir veicinājusi sāncensību labāku modeļu izstrādē, padarījusi pētniecību dziļās mašīnmācīšanās jomā pieejamāku un ļāvusi pētniekiem strādāt pie attēlu izpratnes uzdevumiem pat tad, ja viņiem nav resursu datu iegūšanai un marķēšanai. Tomēr praktiskiem pielietojumiem bieži vien ir nepieciešamas specializētas datu kopas, un tas rada vairākus ar datiem saistītus izaicinājumus, pielietojot dziļo mašīnmācīšanos lietišķām problēmām, piemēram, nepieciešamību savākt datus pietiekamā daudzumā, apkopot un nomarķēt tos, kā arī nepieciešamību risināt jēdzienu novirzes (angl. *concept drift*) problēmu, t.i., izmaiņas pašreizējā klašu sadalījumā salīdzinājumā ar klašu sadalījumu modeļa apmācības

laikā [20].

Lai mazinātu datu nepietiekamības un nelīdzsvarotības problēmas, var izmantot vairākas pieejas, proti:

- **zināšanu pārnese** (angl. *transfer learning*), kuras mērķis ir uzlabot modeļa precizitāti uz mērķa domēna, izmantojot zināšanas, kas iegūtas citā domēnā – avota domēnā [21];
- **precīzā pielāgošana** (angl. *fine-tuning*), kas ietver iepriekš iesaldēto modeļa slāņu atsaldēšanu un to apmācīšanu uz mērķa datu kopas, izmantojot nelielu mācīšanās ātrumu;
- **datu paplašināšana** (angl. *data augmentation*), t.i., papildu informācijas iegūšana modeļa apmācībai no sākotnējiem datiem, mākslīgi palielinot apmācības datu kopu. To var veikt, deformējot attēlus vai vairākkārt tos izmantojot [22];
- **sintētisko datu izmantošana**, tas ir, modeļa apmācība uz mākslīgi generētiem datiem, kas ir līdzīgi mērķa domēna datiem.

2. Roku mazgāšanas kustību klasifikācija

Slikta roku higiēna ir galvenais pret vairākiem medikamentiem rezistentu baktēriju un infekciju izplatības cēlonis klīniskajos apstākļos [23]. Roku higiēnas norāžu ievērošana samazina ar veselības aprūpi saistīto infekciju izplatību [24], tāpēc vadlīniju ievērošana ir vitāli svarīga. Pasaules Veselības organizācija (PVO) piedāvā plaši pieņemtas roku mazgāšanas vadlīnijas [25], kurās ir ietvertas sešas galvenās kustības. Tomēr pētījumi [26, 27, 28] liecina, ka gan sabiedrība kopumā, gan veselības aprūpes darbinieki bieži neievēro šos soļus, tādējādi palielinot infekciju izplatības risku.

Būtiska loma PVO vadlīniju ievērošanā ir roku mazgāšanas novērošanai. Parasti to veic, izmantojot tiešo novērošanu [29, 30, 24]. Automatizētas sistēmas var nodrošināt nepārtrauktu un precīzāku novērošanu, veicot roku mazgāšanas epizožu video ierakstus, klasificējot kustības, sniedzot atgriezenisko saiti lietotājiem un saglabājot datus turpmākai analīzei. Daudzsološa pieeja ir izmantot viedtālrunus vai robežierīces, kā izmaksu ziņā efektīvus šādu sistēmu kodolus, tādējādi vienkāršojot novērošanas sistēmu uzstādīšanu slimnīcās. Nesenos pētījumos [31, 32, 33] tika sasniegti daudzsološi rezultāti, izmantojot datorredzes metodes, piemēram, KNT balstītus klasifikatorus, bet nav skaidrs, vai augstā precizitāte, ko šādas sistēmas sasniedz uz laboratorijas apstākļos savāktām roku mazgāšanas datu kopām, saglabāsies, sistēmām darbojoties reālās pasaules apstākļos. Lai risinātu šo problēmu, šajā nodaļā aprakstītie pētījumi ietvēra eksperimentus ar sarežģītākām reālās pasaules datu kopām. **Galvenā hipotēze** bija, ka zemu resursu patēriņa KNT, t.i., KNT, kas spēj darboties uz mobilajām un robežierīcēm, spēj veiksmīgi, t.i., ar precizitāti, kas pārsniedz hipotētiskā primitīvā klasifikatora precizitāti, klasificēt roku mazgāšanas kustības reālās slimnīcas vidē.

2.1. Datu kopas

Lai risinātu brīvpiekļuves datu trūkumu mašīnmācīšanās balstītu roku mazgāšanas klasifikatoru apmācībai, tika izveidotas divas datu kopas, PSKUS un METC, savācot datus klīniskajā vidē un nomarkējot tos.

2.1.1. *PSKUS datu kopa*

PSKUS datu kopa tika savākta Paula Stradiņa Klīniskajā universitātes slimnīcā, izmantojot pielāgotu lietu interneta sistēmu. Kameras tika uzstādītas deviņās vietās slimnīcā virs izlietnēm un ilgstoši veica roku mazgāšanas epizožu ierakstus. Dažās vietās tika uzstādīta vairāk nekā viena

kamera, lai veiktu ierakstus no dažādiem leņķiem.

Datu kopas marķēšanu veica eksperti infektologi un brīvprātīgie. Datu anotatori vispirms marķēja katru kadru, lai norādītu, vai tajā notiek roku mazgāšana, un pēc tam klasificēja kustības atbilstoši PVO vadlīnijām. Datu kopa satur 3 185 video ar 6 690 anotācijām, kas aptver vairāk nekā 83 000 sekunžu ierakstu. Marķēšana sastāv no septiņām klasēm: sešas roku mazgāšanas kustības tika marķētas atbilstoši PVO vadlīnijām, kā arī tika pievienota papildu klase 0 – “citas kustības”.

Lielāko daļu video marķēja vairāki anotatori; rezultātā anotatoru viedokļi par roku mazgāšanas esamību sakrita 91,23% gadījumū, savukārt anotatoru viedokļi par kustību klasifikāciju, ja tika konstatēta roku mazgāšana, sakrita 90,06% gadījumū.

2.1.2. METC datu kopa

METC datu kopa tika savākta Rīgas Stradiņa universitātes Medicīnas izglītības tehnoloģiju centrā, kā daļa no pētījuma [34] par atgriezeniskās saites sniegšanu lietotājiem, kurā piedalījās 72 veselības aprūpes speciālisti. Datu vākšana notika vienā lokācijā. Katrs dalībnieks veica trīs roku mazgāšanas mēģinājumus ar dažādu atgriezenisko saiti: bez norādēm, ar viedtālruņa lietotnes sniegtajām daļējām norādēm un ar detalizētām norādēm, izmantojot to pašu lietotni.

Roku mazgāšanas epizodes tika ierakstītas, kā arī tas reāllaikā marķēja operators. Datu kopu veido 212 video ar 212 anotācijām, kas aptver 13 870 sekunžu ierakstu.

2.2. Sākotnējie eksperimenti uz PSKUS un METC datu kopām

Sākotnējos eksperimentos es izmantoju MobileNetV2 KNT modeļus. Pirmajā eksperimentā tika iesaldēts bāzes modelis, savukārt pievienotie slāņi – pilnībā savienotais slānis ar 128 neironiem un izvades slānis ar 7 neironiem — tika apmācīti 30 ciklu garumā. Uz PSKUS datu kopas apmācītais modelis sasniedza 16,72% F_1 vērtību, savukārt uz METC datu kopas apmācītais modelis parādīja daudz labāku sniegumu, sasniedzot 49,74% F_1 vērtību.

Otrajā eksperimentā es atsaldēju visus modeļa slāņus un turpināju apmācību vēl 30 ciklu garumā ar samazinātu mācīšanās ātrumu. Uzlabojumi uz PSKUS datu kopas netika sasniegti, jo tika sasniegts 15,52% F_1 vērtību, savukārt rezultāti uz METC datu kopas parādīja būtisku uzlabojumu, F_1 vērtībai pieaugot līdz 63,89%.

Atšķirības starp rezultātiem, kas tika sasniegti uz šīm datu kopām, uzskatāmi demonstrē, cik sarežģīti modelim vispārināt, ja dati iegūti vairākās

vietās, un izmantot vairāku anotatoru radītu mazāk konsekvento marķēšanu PSKUS datu kopā. Pretstatā tam, eksperimenti ar METC datu kopu, kurai bija viens anotators un vairāk kontrolēta vide, uzrādīja labākus rezultātus.

2.3. KNT efektivitātes izpēte uz dažādām datu kopām

Ņemot vērā to, ka sākotnējos eksperimentos KNT neuzrādīja tik augstus rezultātus roku mazgāšanas kustību klasifikācijā, kā citās publikācijās atrodamie rezultāti, es un mani kolēģi veica pētījumu, lai noskaidrotu, vai KNT arhitektūras, kas uzrāda labus rezultātus uz vienkāršākām un salīdzinoši nelielām datu kopām, spēj sasniegt tikpat labus rezultātus uz lielākām un sarežģītākām datu kopām. Papildus PSKUS un METC datu kopām mēs izmantojām arī Kaggle roku mazgāšanas datu kopas [35] publiski pieejamo daļu, kas sastāv no 25 videoierakstiem ar uzsvērti rūpīgu roku mazgāšanu. Mūsu eksperimentos mēs turpinājām izmantot MobileNet tipa klasifikatorus, kuriem nav nepieciešama jaudīga aparatūra, un līdz ar to tie ir piemēroti roku mazgāšanas novērošanas sistēmu izveidei reālās pasaules apstākļos. Mēs iekļāvām mūsu pētījumā arī publikācijās bieži sastopamas vairāklūsmu neironu tīklu arhitektūras un arhitektūras ar rekurentiem elementiem.

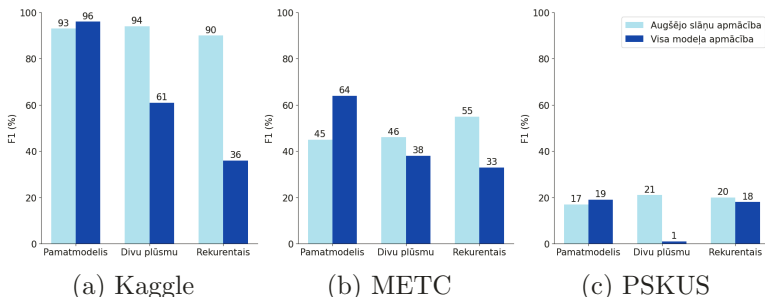
2.3.1. Metodoloģija

Eksperimentos izmantotajam pamatmodelim bija tāda pati arhitektūra, kā sākotnējos eksperimentos. Lai uzlabotu roku mazgāšanas kustību atpazīšanu laikā, tika testēti divi papildu arhitektūras veidi. Pirmais veids bija divu plūsmu tīkls, kurā viens MobileNetV2 modelis apstrādāja RGB ievadi, bet otrs apstrādāja optisko plūsmu, lai uztvertu kustību starp kadriem. Otra arhitektūra bija rekurentais KNT, kuru veidoja pieci paralēli MobileNetV2 modeļi savienoti ar vienu Gated Recurrent Unit (GRU; [36]) moduli ar 256 neironiem.

Pamatmodelis un rekurentais modelis tika izmantoti gan ar augstāk minēto konfigurāciju, gan ar diviem papildu pilnīgi savienotiem slāņiem ar 128 neironiem katrā, kas tika pievienoti pirms izvades slāņa.

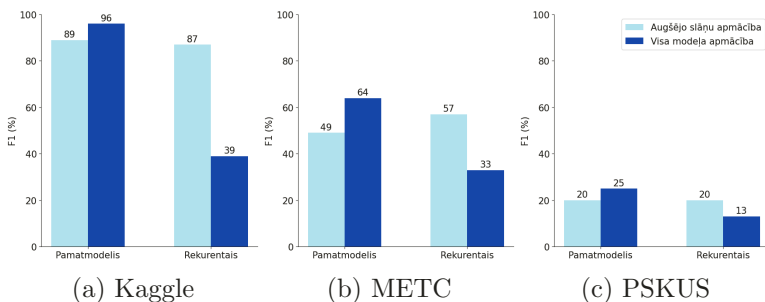
Modeļi tika apmācīti 20 ciklu garumā. Mēs veicām eksperimentus, apmācot gan visu modeli, gan tikai augšējos slāņus. Tika veikti papildus eksperimenti ar zināšanu pārnesei, apmācot modeļus 10 ciklu garumā, lai novērtētu modeļa spēju vispārināt, pārbaudot zināšanu pārnesei no vienkāršākām datu kopām uz sarežģītākām.

2.3.2. Rezultāti



Attēls 2.1.: Dažādu KNT arhitektūru F_1 vērtības.

Galveno eksperimentu rezultāti (Attēls 2.1.) parādīja, ka pamata vienkadra modelis uzrādīja labākos rezultātus gan uz Kaggle (96% F_1 vērtība), gan uz METC (64% F_1 vērtība) datu kopām. Pārsteidzošā kārtā, sarežģītāki modeļi – divu plūsmu tīkls un rekurentais KNT – uzrādīja sliktākus rezultātus. Uz PSKUS datu kopas labākais modelis (divu plūsmu tīkls, kuram tika apmācīti tikai augšējie slāņi) sasniedza vien 21% F_1 vērtību, kas norāda uz modeļu vāju spēju vispārināt uz sarežģītiem reālās pasaules datiem.

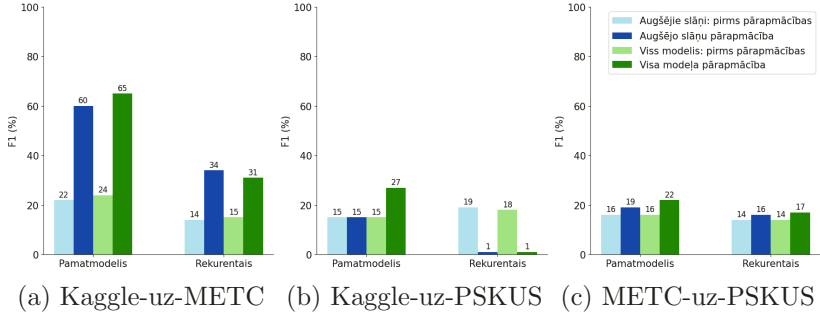


Attēls 2.2.: Dažādu KNT arhitektūru, kuras papildinātas ar diviem pilnībā savienotiem slāņiem, F_1 vērtības.

Eksperimenti ar divu pilnīgi savienotu slāņu pievienošanu (Attēls 2.2.) pamata un rekurentiem modeļiem¹ parādīja līdzīgas tendences: vienkadra modelis joprojām uzrādīja labākos rezultātus uz Kaggle (96% F_1 vērtība)

¹Ņemot vērā to, ka divu plūsmu tīkls nevienā no eksperimentiem neuzrādīja būtiskus uzlabojumus salīdzinājumā ar divām pārējām pieejām, mēs neizmantojām šo arhitektūru turpmākajos eksperimentos.

un METC (64% F_1 vērtība) datu kopām, taču uzlabojumi uz PSKUS datu kopas bija minimāli, labākai F_1 vērtībai sasniedzot tikai 25%.



Attēls 2.3.: Dažādu KNT arhitektūru F_1 vērtības pēc zināšanu pārnese.

Visbeidzot, zināšanu pārnese eksperimentos (Attēls 2.3.) tika sasniegti daļēji panākumi. Modelis ar labāko veiktspēju, Kaggle-uz-METC pamatmodelis, sasniedza 65% F_1 vērtību uz METC datu kopas, taču spēja vispārināt uz PSKUS saglabājās vāja, maksimāli F_1 vērtībai sasniedzot tikai 27%. Tas liecina, ka zināšanas, kas iegūtas uz vienkāršākām datu kopām, netiek efektīvi pārnestas uz sarežģītākām reālās pasaules datu kopām – tādām kā PSKUS.

Kopumā modeļiem, kas uzrādīja labus rezultātus uz vienkāršākās Kaggle datu kopas, bija grūtības ar sarežģītākām, reālai pasaulei atbilstošām datu kopām – METC un jo īpaši PSKUS. Laika dimensijas iekļaušana vai papildu slāņu pievienošana neveicināja precizitātes uzlabojumus, savukārt visu modeļu slāņu apmācība bija noderīga tikai vienkadra modeļiem. Pētījuma rezultāti apstiprina **galveno hipotēzi** visos eksperimentos, izņemot tos, kas veikti uz PSKUS datu kopas, kā arī demonstrē datu kopas sarežģītības nozīmi, vērtējot roku mazgāšanas klasifikācijas precizitāti reālās pasaules apstākļos.

3. Ielu skatu semantiskā segmentācija

Ielu skatu semantiskā segmentācija ir svarīgs uzdevums pašbraucošo auto izveidē, kas ir daudzsološa un strauji augoša tehnoloģija, kā arī aktīva pētījumu joma zinātnē un nozarē. Līdzīgi kā citos vizuālos domēnos, labākie rezultāti ielu skatu semantiskajā segmentācijā ir sasniegti ar KNT – piemēram, modeļiem no populārās DeepLab bibliotēkas [37]. Taču ņemot vērā to, ka pilsētas skatu attēlu iegūšana un marķēšana KNT apmācībai ir dārga un laikietilpīga, publiski pieejamas datu kopas ar segmentācijai marķētiem ielu skatiem, piemēram, CamVid [38], KITTI [39] un Cityscapes [40], parasti ir salīdzinoši nelielas.

Sintētisko datu izmantošana var palīdzēt risināt datu trūkuma problēmu. Pētījumi ir demonstrējuši, ka šim mērķim var izmantot datorspēlēs iegūtos attēlus [41] un speciāli veidotas sintētiskās ielu skatu datu kopas [42]. Taču vienkāršāka alternatīva ir izmantot atvērtā pirmkoda braukšanas simulatorus, piemēram, TORCS [43] vai CARLA [44]. Tomēr joprojām paliek aktuāls jautājums, vai ar šādiem rīkiem ģenerēti sintētiskie dati ir pietiekami augstas kvalitātes, lai tie būtu noderīgi semantiskās segmentācijas modeļu apmācībai.

Šajā nodaļā aprakstītā pētījuma **mērķis** bija uzlabot ielu skatu semantiskās segmentācijas precizitāti, papildinot reālās pasaules attēlu datu kopu ar sintētiskiem datiem. Proti, es izpētīju, vai ir iespējams uzlabot semantiskās segmentācijas precizitāti, izmantojot sintētiskos datus, kas ģenerēti ar CARLA, ko var paveikt salīdzinoši vienkāršā, ātrā un lielā mērā automatizētā veidā. **Galvenā hipotēze** bija, ka reālās pasaules datu papildināšana ar sintētiskiem datiem uzlabos ielu skatu semantiskās segmentācijas precizitāti.

3.1. Datu kopas

Eksperimentos es izmantoju trīs datu kopas: reālās pasaules Cityscapes datu kopu, MICC-SRI datu kopu [45] un CCM (angl. *Cityscapes-CARLA Mixed*) datu kopu, ko izveidoju šim pētījumam.

Cityscapes datu kopā ir 5 000 attēli ar precīzu (pikseļu līmeņa) marķējumu; attēlu izšķirtspēja ir 1024×2048 pikseļi, un tie uzņemti 50 dažādās Eiropas pilsētās. Šim pētījumam es izveidoju pielāgotu datu kopas sadalījumu – 2 685 attēli apmācībai, 290 attēli validācijai un 500 attēli testēšanai – jo oriģinālā testa kopa nav pieejama brīvpieklūvē, lai to varētu izmantot modeļu salīdzinošai vērtēšanai.

MICC-SRI sastāv no 11 913 sintētiskiem attēliem ar 600×800 pikseļu izšķirtspēju un to segmentācijas maskām. Gan attēli, gan maskas ir ģenerēti

ar CARLA. Attēliem ir zems fotoreālistiskuma līmenis.

CCM datu kopa, ko es izveidoju šim pētījumam, sastāv no 2 685 Cityscapes attēliem un 46 935 sintētiskiem attēliem, kurus es ģenerēju ar CARLA. Sintētisko attēlu izšķirtspēja ir 1024×2048 pikseļi, kas atbilst Cityscapes attēlu izšķirtspējai. Lai ģenerētu datus, es veicu simulācijas vairākās pilsētās kartēs, kas bija pieejamas CARLA jaunākajā stabilaajā versijā (v0.9.12). Katrā kartē tika nejauši novietoti 100 transportlīdzekļi un 200 gājēji.

3.2. Metodoloģija

Pirms CNN apmācības es veicu datu pirmapstrādi, pārmarķējot semantiskās segmentācijas maskas un mainot attēlu izmērus, lai datu kopas būtu saderīgas savā starpā. Kā arī es izveidoju vairākus CCM datu kopas sadalījumu veidus — CCM-100, CCM-50 un CCM-25 — papildinot reālās pasaules Cityscapes datu kopu ar pieejamo sintētisko datu attiecīgo procentu.

Pēc datu pirmapstrādes es veicu semantiskās segmentācijas eksperimentus, izmantojot divus KNT modeļus no DeepLabv3 bibliotēkas – MobileNetV2 un Xception-65. Modeļi tika apmācīti, izmantojot DeepLabv3 bibliotēkas noklusējuma iestatījumus. Modeļu validācijai un testēšanai es izmantoju tikai reālās pasaules Cityscapes attēlus.

Eksperimentos ar MICC-SRI un Cityscapes MobileNetV2 modeļi tika apmācīti 1200 ciklu garumā uz katras datu kopas, bet Xception-65 modeļi — 300 ciklu garumā uz katras datu kopas. Eksperimentos ar CCM un Cityscapes MobileNetV2 un Xception-65 modeļi tika apmācīti 200 ciklu garumā uz Cityscapes un CCM-100, CCM-50 un CCM-25 datu kopu sadalījumiem. Apmācība tika veikta uz Rīgas Tehniskās universitātes (RTU) Augstas veiktspējas skaitļošanas centra (angl. *HPC – High-Performance Computing Center*) serveriem un prasīja ≈ 2900 skaitļošanas stundas.

3.3. Rezultāti

Es aprakstu galvenos rezultātus, izmantojot semantiskās segmentācijas standarta metrikas: šķēluma un apvienojuma attiecību (angl. *Intersection over Union* jeb *IoU*) katrai klasei un visu klašu vidējo IoU (angl. *mean Intersection over Union* jeb *mIoU*). Līdzīgi kā citi autori (skat., piem., [40]), es norādu un iekļauju aprēķinos tikai semantiski nozīmīgas klases, izslēdzot tādas klases kā *Cits* vai *Neviens*.

Rezultāti uz Cityscapes un MICC-SRI datu kopām

MobileNetV2 un Xception-65 apmācības rezultāti uz Cityscapes and MICC-SRI ir apkopoti Tabulās 3.1. un 3.2.

Tabula 3.1. Semantiskās segmentācijas precizitātes (IoU) salīdzinājums: MobileNetV2, kas apmācīts uz Cityscapes, salīdzinājumā ar MobileNetV2, kas apmācīts uz ar MICC-SRI paplašinātās Cityscapes.

Klase	Cityscapes	Cityscapes & MICC-SRI
Ceļš	92,66	92,62
Ietve	67,02	66,61
Ēka	86,48	86,18
Žogi un sienas	44,46	43,21
Stabi un ceļa zīmes	57,07	56,72
Apstādījumi	89,52	89,45
Gājēji	76,59	76,54
Transportlīdzekļi	89,65	89,55
mIoU	75,43	75,11

Tabula 3.2. Semantiskās segmentācijas precizitātes (IoU) salīdzinājums: Xception-65, kas apmācīts uz Cityscapes, salīdzinājumā ar Xception-65, kas apmācīts uz ar MICC-SRI paplašinātās Cityscapes.

Klase	Cityscapes	Cityscapes & MICC-SRI
Ceļš	93,69	93,60
Ietve	71,78	72,70
Ēka	88,67	88,30
Žogi un sienas	52,20	49,16
Stabi un ceļa zīmes	63,58	62,52
Apstādījumi	90,75	90,58
Gājēji	81,75	81,39
Transportlīdzekļi	92,29	92,24
mIoU	79,34	78,81

Kā var redzēt, Cityscapes datu kopas paplašināšana ar MICC-SRI attēliem neuzlaboja semantiskās segmentācijas precizitāti: gluži otrādi, gan MobileNetV2, gan Xception-65 modeļi uzrādīja nedaudz labākus rezultātus, kad tie tika apmācīti tikai uz reālās pasaules attēliem, salīdzinot ar apmācību uz paplašinātās datu kopas. Ticamākais skaidrojums modeļu sliktākai precizitātei uz paplašinātās datu kopas ir MICC-SRI datu kopas attēlu zems fotoreālistiskums.

Rezultāti uz Cityscapes un CCM datu kopām

MobileNetV2 un Xception-65 apmācības rezultāti uz Cityscapes un trim CCM sadalījumiem — CCM-100, CCM-50 un CCM-25 — ir apkopotī Tabulās 3.3. un 3.4. Kā var redzēt, abu KNT arhitektūru gadījumā datu kopas paplašināšana ar sintētiskiem datiem uzlaboja semantiskās segmentācijas precizitāti, tādējādi apstiprinot **galveno hipotēzi**.

Tabula 3.3. Semantiskās segmentācijas precizitātes (IoU) salīdzinājums: MobileNetV2 modeļi, kas apmācīti uz Cityscapes, CCM-100, CCM-50 un CCM-25 datu kopām.

Class	Cityscapes	CCM-100	CCM-50	CCM-25
Ēka	75,54	79,39	80,17	79,18
Žogs	00,02	21,49	24,47	17,82
Gājējs / Braucējs	69,23	67,94	68,92	69,38
Stabs	10,48	38,51	38,54	36,81
Ceļš	88,57	88,46	89,72	89,15
Ietve	54,51	57,24	59,79	58,62
Apstādījumi	83,90	86,14	86,41	85,54
Transportlīdzeklis	82,20	82,72	82,72	82,78
Siena	0,00	19,90	23,81	15,95
Ceļa zīme	0,00	35,42	34,10	25,30
Debess	82,80	85,32	85,83	85,85
Luksofors	0,00	21,32	15,38	00,13
Ūdens / Zeme	33,61	37,50	38,73	38,20
mIoU	44,68	55,49	56,05	52,67

Tabula 3.4. Semantiskās segmentācijas precizitātes (IoU) salīdzinājums: Xception-65 modeļi, kas apmācīti uz Cityscapes, CCM-100, CCM-50 un CCM-25 datu kopām.

Class	Cityscapes	CCM-100	CCM-50	CCM-25
Ēka	84,94	85,10	85,08	85,62
Žogs	37,20	40,19	40,19	43,44
Gājējs / Braucējs	78,08	76,42	76,94	77,92
Stabs	45,08	48,50	48,75	49,26
Ceļš	92,31	91,82	91,43	91,85
Ietve	65,80	67,21	67,09	69,88
Apstādījumi	87,71	87,00	87,59	87,85
Transportlīdzeklis	89,86	88,82	89,63	89,63
Siena	23,29	28,88	27,69	31,63
Ceļa zīme	44,42	50,89	55,83	56,14

Tabula 3.4. – turpinājums.

Debess	85,13	88,79	89,70	90,34
Luksofors	0,00	43,64	42,19	44,13
Ūdens / Zeme	46,86	39,58	35,62	44,22
mIoU	57,25	63,14	63,87	64,46

Jāatzīmē, ka labākie MobileNetV2 un Xception-65 modeļi nebija tie, kas tika apmācīti uz CCM sadalījumiem ar lielāko sintētisko datu apjomu: labākais MobileNetV2 modelis tika apmācīts uz CCM-50, bet labākais Xception-65 modelis – uz CCM-25. Tas norāda uz to, ka lielāka sintētisko datu apjoma izmantošana datu kopas paplašināšanai ne vienmēr sniedz labākus rezultātus, nekā paplašināšana ar mazāku sintētisko datu apjomu. Turklāt šis pētījums arī parādīja, ka sintētisko datu ģenerēšanas sistēmas izveidei nav jābūt dārgai vai sarežģītai, jo izmantojot CARLA standarta konfigurāciju, bija iespējams ātri ģenerēt lielu daudzumu sintētisko datu.

4. Objektu atpazīšana groza izkraušanas uzdevuma veikšanai

Vienpakāpes objektu detektori, piemēram, YOLO [46], bieži tiek izmantoti rūpnieciskajās robotu sistēmās objektu satveršanai, jo tie var atpazīt objektus gan ātri, gan precīzi. Tomēr šādu modeļu apmācībai nepieciešamas lielas datu kopas. Simulāciju un sintētisko datu izmantošana ir daudzsološs risinājums datu pieejamības problēmai uztveres sistēmu izveidē robotikā. Tomēr modeļiem, kas apmācīti tikai uz sintētiskajiem datiem, bieži samazinās efektivitāte reālās pasaules uzdevumos atšķirību starp mākslīgo un reālo domēnu dēļ [47]. Lai uzlabotu sintētisko datu reālistiskumu, var izmantot domēna adaptācijas tehnikas, piemēram, dažādus ģeneratīvus pretinieku tīklus (angl. *generative adversarial networks* jeb *GAN*; [48]).

Šajā nodaļā aplūkotā pētījuma **galvenais mērķis** bija uzlabot plastmasas pudeļu ar augstu redzamību (t.i., kaudzes augšpusē esošo) atpazīšanas precizitāti ar YOLOv5 objektu detektoru [17], lai robota roka labāk varētu tās satvert. **Galvenā hipotēze** bija, ka sintētisko plastmasas pudeļu attēlu fotoreālistiskuma uzlabošana ar GAN pirms YOLOv5 apmācības uz tiem uzlabos objektu atpazīšanas precizitāti salīdzinājumā ar apmācību uz oriģināliem sintētiskajiem attēliem.

4.1. Datu kopas

Pētījumā tika izmantotas vairākas datu kopas, proti, reālās pasaules pudeļu attēlu datu kopa, sintētiskā pudeļu attēlu datu kopa un vairākas datu kopas, kas izveidotas, uzlabojot sintētisko datu kopu ar GAN.

Reālās pasaules datu kopa tika iegūta pētījumā [49], nejauši izvietojot pudeles plastmasas konteinerā un uzņemot attēlus ar dažādām pudeļu pozīcijām, kameras eksponēcijas laikiem un apgaismojumu. Tajā ir 2 060 manuāli marķēti attēli, no kuriem 1 760 tika izmantoti GAN apmācībai, bet 300 tika izmantoti objektu detektoru testēšanai.

Sintētiskā datu kopa tika ģenerēta pētījumā [49], izmantojot Blender fizikas simulācijas dzinēju. Tajā ir 8 800 augstas izšķirtspējas attēli ar pudelēm, kas nejaušā secībā tika novietotas simulētā kastē, mainot apgaismojuma intensitāti un skata leņķus. Par satveramām tika uzskatītas pudeles ar vairāk nekā 60% redzamību.

Lai izveidotu datu kopas ar uzlabotu fotoreālistiskumu, tika izmantots CycleGAN [50], jo tas, ka tam nav nepieciešams, lai reālie un sintētiskie attēli veidotu pārus, atviegloja uzdevumu. Izmēģinot dažādas pieejas, tika izveidotas vairākas datu kopas, proti:

- *Baseline* CycleGAN – izmantojot CycleGAN ar noklusējuma parametriem;
- *Augmented* CycleGAN – izmantojot tikai attēla centra daļu un pievienojot pirmapstrādes funkcijas, piemēram, nejaušu kontrastu, spilgtumu, toni un piesātinājumu, lai samazinātu pārmācīšanās risku;
- *Augmented noise* CycleGAN – pievienojot Gausa troksni diskriminatora ievadei, lai tam būtu grūtāk novērtēt attēlus, un līdz ar to ļaujot ģeneratoram ilgāk mācīties;
- *Resized convolution* CycleGAN – aizstājot transponētās konvolūcijas operācijas [51] GAN dekoderā ar izmēra maiņas operāciju un tai sekojošu konvolūciju, lai novērstu “šaha dēļa” (angl. *checkerboard*) artefaktu [52] rašanos;
- *Resized transpose* CycleGAN – piemērojot izmēra maiņas operāciju un tai sekojošu konvolūciju visiem slāņiem, izņemot pēdējo, kurā transponētā konvolūcija tika saglabāta. Tas tika veikts, lai novērstu izplūdumu, kas parādījās *Resized convolution* CycleGAN datu kopas attēlos.

Pēc GAN uzlaboto datu kopu izvērtēšanas gan vizuāli, gan aprēķinot FID vērtējumu [53], objektu atpazīšanas eksperimentiem tika izvēlētas *Augmented noise* CycleGAN un *Resized transpose* CycleGAN datu kopas. Šīs datu kopas tika izmantotas ar 1024×768 pikseļu izšķirtspēju, jo attēli ar šādu izšķirtspēju tika labāk novērtēti nekā attēli ar mazāku 256×256 pikseļu izšķirtspēju, kas tika izmantoti sākotnēji.

4.2. Metodoloģija

Es veicu objektu detekcijas eksperimentus, izmantojot YOLOv5 no Ultralytics bibliotēkas. Eksperimenti tika veikti uz trim datu kopām: oriģināliem sintētiskiem attēliem, *Augmented noise* CycleGAN 1024×768 datu kopas un *Resized transpose* CycleGAN 1024×768 datu kopas. Es izmantoju trīs dažāda izmēra YOLOv5 modeļus — Small (7.2M parametru), Medium (21.2M parametru) un Extra Large (86.7M parametru). Modeļi tika apmācīti 300 ciklu garumā un testēti uz datu kopas, kurā bija 300 reālās pasaules attēli.

4.3. Rezultāti

Objektu atpazīšanas eksperimentu rezultāti ir apkopoti Tabulā 4.1., izmantojot standarta metrikas: precizitāti, pārklājumu un objektu robežtaisnstūru noteikšanas vidējo precizitāti (angl. *mean average precision* jeb *mAP*). Pēdējā metrika tika aprēķināta gan ar pārklājuma sliekšni 0,5,

gan kā vidējā vērtība dažādiem pārklājuma sliekšņiem – no 0,5 līdz 0,95 ar soli 0,05.

Tabula 4.1. Objektu atpazīšanas eksperimentu rezultāti ar YOLOv5, izmantojot oriģinālos sintētiskos datus un sintētiskos datus, kas uzlaboti ar CycleGAN.

Modelis	Datu kopa	Prec.	Pārkl.	mAP (sliexsnis 0,5)	mAP (vidējais sliexšņiem [0,5 : 0,05 : 0,95])
YOLOv5 Small	Oriģinālā sintētiskā	68,9	89,4	74,2	44,7
	<i>Resized transpose</i> 1024 × 768	61,5	83,8	63,2	24,5
	<i>Augmented noise</i> 1024 × 768	73,1	90,5	77,7	48,0
YOLOv5 Medium	Oriģinālā sintētiskā	69,3	80,2	72,0	42,0
	<i>Resized transpose</i> 1024 × 768	58,1	78,2	58,9	20,4
	<i>Augmented noise</i> 1024 × 768	71,2	91,0	75,3	45,5
YOLOv5 Extra Large	Oriģinālā sintētiskā	71,8	78,5	75,0	41,3
	<i>Resized transpose</i> 1024 × 768	68,3	86,7	71,5	33,9
	<i>Augmented noise</i> 1024 × 768	72,1	87,2	76,1	46,1

Prec.: precīzumspēja. **Pārkl.:** pārklājums.

Kaut arī modeļi, kas tika apmācīti uz *Resized transpose* 1024 × 768 CycleGAN datu kopas, kopumā uzrādīja sliktākus rezultātus nekā modeļi, kas tika apmācīti uz oriģinālajiem sintētiskajiem datiem, modeļi, kas tika apmācīti uz *Augmented noise* 1024 × 768 CycleGAN datu kopas, pārspēja uz oriģinālajiem sintētiskajiem datiem apmācītos modeļus visās metrikās, tādējādi apstiprinot **galveno hipotēzi**. Jāatzīmē, ka lielākie modeļi ne vienmēr pārspēja mazākos, kas var nozīmēt, ka lielākie modeļi, iespējams, ir pārmācījušies uz salīdzinoši neliela apjoma apmācību datu kopām.

5. Attēlu klasifikācija orgānu uz čipa audzēšanas novērošanai

Orgāni uz čipa ir daudzsoļa biomedicīnas tehnoloģija, kas apvieno audu inženieriju un mikrofluidiku, lai imitētu svarīgus cilvēka fizioloģijas aspektus ar mērķi atveidot kādu cilvēka orgānu iekšējo vidi ārpus organisma. Pašlaik iekārtu orgānu uz čipa audzēšanai novērošanu veic cilvēki, taču būtu lietderīgi automatizēt šo procesu. Viens no galvenajiem novērošanas aspektiem ir orgānu uz čipa šūnu paraugu kvalitātes novērtēšana, tāpēc šajā nodaļā aplūkotā pētījuma **mērķis** bija izstrādāt orgānu uz čipa mikroskopijas attēlu klasifikatoru. Ņemot vērā KNT panākumus vairākos biomedicīnas uzdevumos [54, 55, 56], es izmantoju šim nolūkam EfficientNet-B7 un MobileNetV3Large un izvirzīju šādu hipotēzi:

Pirmā hipotēze: KNT balstīts klasifikators sasniedz labāku precizitāti uz orgānu uz čipa reālās pasaules mikroskopijas attēlu datu kopas nekā hipotētiskais primitīvais klasifikators.

Galvenais izaicinājums DNT balstītu biomedicīnas attēlu klasifikatoru izveidē ir datu pieejamība, jo biomedicīnas datu kopas bieži ir nelielas un nesabalansētas. Sintētisko datu izmantošana ir daudzsoļa pieeja šīs problēmas risināšanai. Šajā nodaļā aprakstītajā pētījumā es ģenerēju sintētiskos datus ar Stable Diffusion [57] latentās difūzijas modeli, kas demonstrējis teicamas spējas dažāda veida attēlu radīšanā. Eksperimentiem ar sintētiskiem attēliem tika izvirzīta sekojoša hipotēze:

Otrā hipotēze: Klasifikācijas precizitāte uz orgānu uz čipa reālās pasaules mikroskopijas attēlu datu kopas uzlabojas, ja KNT balstīts klasifikators tiek apmācīts uz datu kopas, kas paplašināta ar sintētiskiem datiem, kas ģenerēti ar Stable Diffusion modeli, nevis tikai uz reālās pasaules attēlu datu kopas.

5.1. Eksperimenti uz sākotnējās orgānu uz čipa attēlu datu kopas

5.1.1. *Dati un metodoloģija*

Sākotnējo orgānu uz čipa attēlu datu kopu veidoja 822 šūnu attēli no dažādām šūnu līnijām: Caco-2 un HUVEC, kas atveido zarnas, A549 un HPMEC, kas atveido plaušu vēzi, un HSAEC, kas atveido plaušas. Attēlu marķēšanu veica pieredzējuši šūnu biologi; rezultātā, pamatojoties uz šūnu morfoloģiju un blīvumu, attēli tika iedalīti trijās klasēs: “labs” (500 attēli), “pieņemams” (212 attēli) un “slikts” (110 attēli).

Lai paplašinātu sākotnējo orgānu uz čipa attēlu datu kopu, es veicu Stable Diffusion precīzo pielāgošanu, izmantojot LoRA [58] – efektīvu metodi, kuras pielietošanai ir nepieciešami vien daži desmiti attēlu – un generējot divas sintētiskās datu kopas ar LoRA svaru vērtībām 1,0 un 0,8.

Sākotnējos eksperimentos es izmantoju EfficientNet-B7 KNT modeli. Apmācības laikā bāzes modeļa svāri tika iesaldēti, savukārt divi pievienotie slāņi – Batch Normalization slānis un izvades slānis ar 3 neironiem – tika apmācīti 30 ciklu garumā.

5.1.2. Rezultāti

Eksperimentu rezultāti uz sākotnējās datu kopas ir parādīti Tabulā 5.1.

Tabula 5.1. EfficientNet-B7 precizitāte pēc apmācības uz sākotnējās datu kopas, kas paplašināta ar sintētiskajiem datiem, ģenerētiem ar Stable Diffusion modeli pēc tā precīzās pielāgošanas ar LoRA.

Datu kopa	LoRA svars 1,0	LoRA svars 0,8
Tikai sint. (100%)	61,4	62,1
Reālā un 100% sint.	69,9	70,1
Reālā un 75% sint.	69,6	70,7
Reālā un 50% sint.	71,0	69,3
Reālā un 25% sint.	70,7	70,4
Reālā un 10% sint.	72,1	71,8
Tikai reālā	72,9	

Bāzes modelis sasniedza precizitāti 72,9%, kas ir labāk nekā hipotētiskais primitīvais klasifikators (60,8%). Taču datu kopas paplašināšana ar sintētiskajiem attēliem noveda pie precizitātes samazināšanās, lielākam sintētisko datu apjomam izraisot sliktākus rezultātus.

5.2. Eksperimenti uz orgānu uz čipa gala datu kopas

5.2.1. Dati un metodoloģija

Orgānu uz čipa attēlu gala datu kopa [59] sastāv no 3 072 attēliem. Tā iekļauj sevī sākotnējo datu kopu, bet arī atšķiras no tās divos būtiskos aspektos: pirmkārt, pamatojoties uz bioloģijas ekspertu ieteikumiem, trīs klašu marķējums tika aizstāts ar skaidrāku bināro klasifikāciju – “labs” un “slikts”; otrkārt, tajā tika iekļauta vēl viena šūnu līnija – NHBE bronhu epitēlija šūnas.

Attēlu sadalījums pa klasēm ir sekojošs: “labs” — 1 727 attēli, “slikts” – 1 345 attēli.

Lai paplašinātu orgānu uz čipa attēlu gala datu kopu ar sintētiskiem datiem, tika izmantotas vairākas metodes, proti:

- **attēla pārveidošana citā attēla** (angl. *image-to-image translation*), kas nozīmē tādu ievades attēla pārveidošanu par izvades attēlu, ka tas gan saglabā dažas iezīmes no sākotnējā attēla, gan iegūst jaunas iezīmes;
- **selektīva attēla aizpildīšana ar maskām** (angl. *inpainting with masks*), kas ietver daļēju ievades attēla pārveidi, piemērojot masku, lai norādītu, kuras attēla daļas jāmaina;
- **attēlu interpolācija**, izveidojot starpattēlu starp diviem ievadattēliem;
- **precīzā pielāgošana ar LoRA**.

Klasifikācijas eksperimentos tika izmantoti EfficientNet-B7 un MobileNetV3Large. EfficientNet-B7 apmācība tika veikta divos posmos: vispirms modelis tika apmācīts 30 ciklu garumā ar iesaldētiem bāzes modeļa svāriem, kam sekoja precīzā pielāgošana 30 ciklu garumā, apmācot augšējos 20 slāņus. MobileNetV3Large apmācība tika veikta līdzīgā veidā, bet otrajā posmā tika apmācīti 15, nevis 20 augšējie slāņi.

5.2.2. Rezultāti

Eksperimentu rezultāti uz gala datu kopas ir parādīti Tabulā 5.2.

Tabula 5.2. EfficientNet-B7 un MobileNetV3Large precizitāte pēc apmācības uz gala datu kopas, kas paplašināta ar sintētiskajiem datiem, ģenerētiem ar attēla pārveidošanu citā attēlā (**img2img**), selektīvo attēla aizpildīšanu ar maskām (**inpaint**), interpolāciju (**interpol**) un Stable Diffusion pēc precīzās pielāgošanas ar LoRA.

Modelis	Datu kopa	img2img	inpaint	interpol	LoRA
EffNet-B7	Reālā un 100% sint.	85	84	80	82
	Reālā un 75% sint.	82	83	81	78
	Reālā un 50% sint.	83	82	80	83
	Reālā un 25% sint.	83	80	81	81
	Tikai reālā			83	
MobNetV3L	Reālā un 100% sint.	81	82	79	76
	Reālā un 75% sint.	81	82	79	77
	Reālā un 50% sint.	81	81	80	80
	Reālā un 25% sint.	82	80	82	80
	Tikai reālā			81	

Pēc apmācības uz reālās pasaules datu kopas bez paplašināšanas ar sintētiskiem datiem, EfficientNet-B7 sasniedza 83% precizitāti, savukārt

MobileNetV3Large sasniedza 81% precizitāti. Abi rezultāti ir labāki nekā hipotētiskā primitīvā klasifikatora precizitāte, kas ir vienāda ar lielākās klases izmēru – 56%; tādējādi eksperimenti apstiprināja **pirmo hipotēzi**.

Eksperimenti ar datu kopas paplašināšanu ar sintētiskiem attēliem parādīja, ka paplašināšana ar attēla pārveidošanas citā attēlā metodi var uzlabot EfficientNet-B7 precizitāti līdz 85% (t.i., par 2%), savukārt MobileNetV3Large gadījumā vairākas metodes — attēla pārveidošana citā attēlā, selektīva attēla aizpildīšana ar maskām un interpolācija — var uzlabot šī modeļa precizitāti līdz 82% (t.i., par 1%). Tādējādi eksperimenti apstiprināja **otro hipotēzi**.

Secinājumi

Promocijas darbā tika aplūkoti datorredzes metožu pielietojumi trim attēlu izpratnes pamatzdevumiem: attēlu klasifikācijai, objektu atpazīšanai un semantiskai segmentācijai. Es izmantoju datorredzes metodes, lai risinātu šādas prakstiskas problēmas:

- roku mazgāšanas kustību klasifikācija klīniskajā vidē, lai automatizētu medicīnas personāla roku higiēnas ievērošanas novērošanu (2. nodaļa);
- ielu skatu semantiskā segmentācija, lai uzlabotu pašbraucošo automašīnu uztveres moduļus (3. nodaļa);
- plastmasas pudeļu atpazīšana satveršanai ar robota roku, lai automatizētu ražošanas līniju (4. nodaļa);
- mikroskopijas attēlu klasifikācija, lai automatizētu orgānu uz čipa audzēšanas novērošanu (5. nodaļa).

Šiem uzdevumiem es pielietoju KNT, izmantojot modeļus ar svariem, kas tika iepriekš apmācīti uz lielām datu kopām, piemēram, ImageNet [60] un MS COCO [61]. Es izmantoju zināšanu pārnēsi un precīzo pielāgošanu, lai pielāgotu modeļus mazākām datu kopām, kas tika izmantotas manos pētījumos. Dažiem uzdevumiem es varēju izmantot piemērotas publiski pieejamas datu kopas: Cityscapes datu kopai bija būtiska loma manā darbā pie ielu skatu semantiskās segmentācijas uzlabošanas, savukārt Kaggle datu kopa ļāva man izveidot pamatu KNT balstītu roku mazgāšanas kustību klasifikatoru efektivitātes novērtēšanai. Kā arī es izmantoju atvērtā pirmkoda resursus, piemēram, CARLA un Stable Diffusion, lai ģenerētu sintētiskus datus reālās pasaules datu kopu papildināšanai. Tomēr datu iegūšana un marķēšana joprojām bija viens no lielākajiem izaicinājumiem, it īpaši strādājot pie roku mazgāšanas kustību atpazīšanas uzdevuma, jo tam nebija publiski pieejamu pietiekami lielu datu kopu, kā arī strādājot pie orgānu uz čipa attēlu klasifikatora izveides, jo pieejamo datu apjoms bija salīdzinoši neliels.

Neskatoties uz izaicinājumiem, uzskatu, ka šī darba pamatu veidojošo pētījumu rezultāti ir bijuši veiksmīgi, proti:

- Roku mazgāšanas kustību klasifikācijas pētījumos tika sasniegti teicami rezultāti – 96% F_1 vērtība uz Kaggle datu kopas un apmērienoši rezultāti – 64% F_1 vērtība uz METC datu kopas. Lai gan neviens modelis nesasniedza labu rezultātu uz sarežģītās un trokšņainās PSKUS datu kopas, minētā datu kopa ir vērtīgs resurss turpmākiem pētījumiem par roku mazgāšanas kustību klasifikāciju.

- Pētījumā par ielu skatu semantiskās segmentācijas precizitātes uzlabošanu Cityscapes papildināšana ar CARLA ģenerētiem datiem ļāva sasniegt augstāku MobileNetV2 un Xception modeļu precizitāti nekā modeļiem, kas tika apmācīti tikai uz Cityscapes.
- Pētījumā par satveramo pudeļu atpazīšanu populārais objektu detektors YOLOv5 demonstrēja, ka tas var efektīvi – labākais modelis sasniedza 77,7% mAP ar sliekšni 0,5 – atpazīt pudeles ar noteiktu redzamību kaudzē ar līdzīgiem objektiem.
- Pētījumā par orgānu uz čipa mikroskopijas attēlu klasifikāciju labākais EfficientNet-B7 modelis sasniedza 85% precizitāti, savukārt labākais MobileNetV3Large modelis sasniedza 82% precizitāti. Ņemot vērā, ka šie modeļi tika apmācīti uz datu kopas, kas tika paplašināta ar Stable Diffusion ģenerētiem attēliem, šis pētījums arī sekmēja izpēti par biomedicīnisko attēlu ģenerēšanu DNT apmācībai ar lieliem ģeneratīviem modeļiem.

Šie rezultāti rāda, ka **darba mērķis** – piedāvāt efektīvus risinājumus lietišķiem attēlu izpratnes uzdevumiem – tika sasniegts visos uzdevumos, izņemot klasifikācijas eksperimentus uz PSKUS datu kopas. **Darba zinātniskā novitāte** ir saistīta ar efektīvu, iepriekš neizpētītu risinājumu piedāvāšanu jaunām datu kopām, kā arī ar metožu uzlabošanu darbam ar jau zināmām datu kopām, izmantojot KNT gan ar reālās pasaules, gan ar sintētiskajiem datiem.

Darba ietvaros veikto pētījumu rezultāti ir publicēti sešos zinātniskajos rakstos (un paredzēts publicēt vēl vismaz divus rakstus), kas indeksēti *Elsevier Scopus* un/vai *Web of Science* datubāzēs un divos zinātniskos izdevumos, kas nav indeksēti šajās datubāzēs, kā arī prezentēti četrās konferencēs. Pastāv arī daudz iespēju turpmākam darbam. Galvenais mērķis tuvākai nākotnei ir izmantot šī darba laikā iegūtās atziņas un zināšanas, lai risinātu attēlu izpratnes uzdevumus citos pašlaik notiekošos pētnieciskos projektos. Lielais mērķis tālākai, taču, manuprāt, paredzamai nākotnei ir piedalīties tādu modeļu izstrādē, kas patiešām spēj *saprast* attēlus.

Bibliogrāfija

- [1] B. Jähne, ed., *Computer Vision and Applications: A Guide for Students and Practitioners*. Elsevier, 2000.
- [2] I. Pitas, *Digital Image processing Algorithms and Applications*. John Wiley & Sons, 2000.
- [3] Y.-J. Zhang and Y.-J. Zhang, “Image engineering,” in *Handbook of Image Engineering*, pp. 55–83, Springer, 2021.
- [4] S. Z. Li, *Markov Random Field Modeling in Computer Vision*. Springer Science & Business Media, 2012.
- [5] Z. Zou, K. Chen, Z. Shi, Y. Guo, and J. Ye, “Object detection in 20 years: A survey,” *Proceedings of the IEEE*, 2023.
- [6] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, “Image segmentation using deep learning: A survey,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 44, no. 7, pp. 3523–3542, 2021.
- [7] N. O’Mahony *et al.*, “Deep learning vs. traditional computer vision,” in *Advances in Computer Vision: Proceedings of the 2019 Computer Vision Conference (CVC), Volume 1*, pp. 128–144, Springer, 2020.
- [8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Advances in neural information processing systems*, vol. 25, pp. 1097–1105, 2012.
- [9] O. Russakovsky *et al.*, “ImageNet large scale visual recognition challenge,” *International journal of computer vision*, vol. 115, pp. 211–252, 2015.
- [10] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, “Deep learning for visual understanding: A review,” *Neurocomputing*, vol. 187, pp. 27–48, 2016.
- [11] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [12] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, “MobileNetV2: Inverted residuals and linear bottlenecks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4510–4520, 2018.
- [13] A. Howard *et al.*, “Searching for MobileNetV3,” in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 1314–1324, 2019.
- [14] M. Tan and Q. Le, “EfficientNet: Rethinking model scaling for convolutional neural networks,” in *International conference on machine learning*, pp. 6105–6114, PMLR, 2019.
- [15] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, “Rethinking atrous convolution for semantic image segmentation,” *arXiv:1706.05587*, 2017.
- [16] F. Chollet, “Xception: Deep learning with depthwise separable convolutions,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1251–1258, 2017.
- [17] G. Jocher, “YOLOv5 by Ultralytics.” [Online], 2020. Available: <https://github.com/ultralytics/yolov5>. Accessed 7 August 2024.

- [18] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.
- [19] A. Glassner, *Deep learning: A visual approach*. No Starch Press, 2021.
- [20] S. Priya and R. A. Uthra, “Deep learning framework for handling concept drift and class imbalanced complex decision-making on streaming data,” *Complex & Intelligent Systems*, vol. 9, no. 4, pp. 3499–3515, 2023.
- [21] F. Zhuang *et al.*, “A comprehensive survey on transfer learning,” *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, 2020.
- [22] C. Shorten and T. M. Khoshgoftaar, “A survey on image data augmentation for deep learning,” *Journal of Big Data*, vol. 6, no. 1, pp. 1–48, 2019.
- [23] S. L. Barnes, D. J. Morgan, A. D. Harris, P. C. Carling, and K. A. Thom, “Preventing the transmission of multidrug-resistant organisms: Modeling the relative importance of hand hygiene and environmental cleaning interventions,” *Infection Control & Hospital Epidemiology*, vol. 35, no. 9, pp. 1156–1162, 2014.
- [24] K. J. McKay, R. Z. Shaban, and P. Ferguson, “Hand hygiene compliance monitoring: Do video-based technologies offer opportunities for the future?,” *Infection, Disease & Health*, vol. 25, no. 2, pp. 92–100, 2020.
- [25] World Health Organization, *WHO guidelines on hand hygiene in health care*. World Health Organization, 2009.
- [26] D. J. Gould, D. Moralejo, N. Drey, J. H. Chudleigh, and M. Taljaard, “Interventions to improve hand hygiene compliance in patient care,” *Cochrane database of systematic reviews*, no. 9, 2017.
- [27] N. Masroor, M. Doll, M. Stevens, and G. Bearman, “Approaches to hand hygiene monitoring: From low to high technology approaches,” *International Journal of Infectious Diseases*, vol. 65, pp. 101–104, 2017.
- [28] B. C. Knepper, A. M. Miller, and H. L. Young, “Impact of an automated hand hygiene monitoring system combined with a performance improvement intervention on hospital-acquired infections,” *Infection Control & Hospital Epidemiology*, vol. 41, no. 8, pp. 931–937, 2020.
- [29] R. T. Ellison III, C. M. Barysaukas, E. A. Rundensteiner, D. Wang, and B. Barton, “A prospective controlled trial of an electronic hand hygiene reminder system,” in *Open Forum Infectious Diseases*, vol. 2, p. ofv121, Oxford University Press, 2015.
- [30] J. M. Boyce *et al.*, “Impact of an automated hand hygiene monitoring system and additional promotional activities on hand hygiene performance rates and healthcare-associated infections,” *Infection Control & Hospital Epidemiology*, vol. 40, no. 7, pp. 741–747, 2019.
- [31] S. Yeung *et al.*, “Vision-based hand hygiene monitoring in hospitals,” *AMIA*, 2016.
- [32] G. Li *et al.*, “Hand gesture recognition based on convolution neural network,” *Cluster Computing*, vol. 22, no. 2, pp. 2719–2729, 2019.
- [33] A. Nagaraj, M. Sood, C. Sureka, and G. Srinivasa, “Real-time action recognition for fine-grained actions and the hand wash dataset,” *arXiv:2210.07400*, 2022.
- [34] O. Zemlanuhina *et al.*, “Influence of different types of real-time feedback on hand washing quality assessed with neural networks/simulated neural networks,” in *SHS Web of Conferences*, vol. 131, pp. 1–13, EDP Sciences, 2022.

- [35] “Sample: Kaggle Hand Wash Dataset.” [Online], 2019. Available: <https://www.kaggle.com/realtimear/hand-wash-dataset>. Accessed 18 February 2024.
- [36] K. Cho, B. Van Merriënboer, D. Bahdanau, and Y. Bengio, “On the properties of neural machine translation: Encoder-decoder approaches,” *arXiv:1409.1259*, 2014.
- [37] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, “Semantic image segmentation with deep convolutional nets and fully connected CRFs,” *arXiv:1412.7062*, 2014.
- [38] G. J. Brostow, J. Fauqueur, and R. Cipolla, “Semantic object classes in video: A high-definition ground truth database,” *Pattern Recognition Letters*, vol. 30, no. 2, pp. 88–97, 2009.
- [39] H. Abu Alhaja, S. K. Mustikovela, L. Mescheder, A. Geiger, and C. Rother, “Augmented reality meets computer vision: Efficient data generation for urban driving scenes,” *International Journal of Computer Vision*, vol. 126, pp. 961–972, 2018.
- [40] M. Cordts *et al.*, “The Cityscapes dataset for semantic urban scene understanding,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3213–3223, 2016.
- [41] S. R. Richter, V. Vineet, S. Roth, and V. Koltun, “Playing for data: Ground truth from computer games,” in *European conference on computer vision*, pp. 102–118, Springer, 2016.
- [42] M. Hahner, D. Dai, C. Sakaridis, J.-N. Zaech, and L. Van Gool, “Semantic understanding of foggy scenes with purely synthetic data,” in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, pp. 3675–3681, IEEE, 2019.
- [43] B. Wymann *et al.*, “TORCS: The open racing car simulator.” [Online], 2015. Available: <https://www.cse.chalmers.se/~chrdimi/papers/torcs.pdf>. Accessed 29 June 2024.
- [44] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, “CARLA: An open urban driving simulator,” in *Conference on robot learning*, pp. 1–16, PMLR, 2017.
- [45] L. Berlincioni, F. Becattini, L. Galteri, L. Seidenari, and A. Del Bimbo, “Road layout understanding by generative adversarial inpainting,” in *Inpainting and Denoising Challenges*, pp. 111–128, Springer, 2019.
- [46] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779–788, 2016.
- [47] N. Jakobi, P. Husbands, and I. Harvey, “Noise and the reality gap: The use of simulation in evolutionary robotics,” in *Advances in Artificial Life: Third European Conference on Artificial Life*, pp. 704–720, Springer, 1995.
- [48] I. Goodfellow *et al.*, “Generative adversarial nets,” *Advances in neural information processing systems*, vol. 27, 2014.
- [49] J. Arents *et al.*, “Synthetic data of randomly piled, similar objects for deep learning-based object detection,” in *International Conference on Image Analysis and Processing*, pp. 706–717, Springer, 2022.
- [50] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” in *Proceedings of the IEEE international conference on computer vision*, pp. 2223–2232, 2017.

- [51] W. Shi *et al.*, “Is the deconvolution layer the same as a convolutional layer?,” *arXiv:1609.07009*, 2016.
- [52] A. Odena, V. Dumoulin, and C. Olah, “Deconvolution and checkerboard artifacts,” *Distill*, vol. 1, no. 10, 2016.
- [53] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, “GANs trained by a two time-scale update rule converge to a local Nash equilibrium,” *Advances in neural information processing systems*, vol. 30, 2017.
- [54] S. A. Ajagbe, K. A. Amuda, M. A. Oladipupo, F. A. Oluwaseyi, and K. I. Okesola, “Multi-classification of Alzheimer disease on magnetic resonance images (MRI) using deep convolutional neural network (DCNN) approaches,” *International Journal of Advanced Computer Research*, vol. 11, no. 53, p. 51, 2021.
- [55] A. Iqbal, M. Sharif, M. A. Khan, W. Nisar, and M. Alhaisoni, “FF-UNet: a U-shaped deep convolutional neural network for multimodal biomedical image segmentation,” *Cognitive Computation*, vol. 14, no. 4, pp. 1287–1302, 2022.
- [56] S. Sharma *et al.*, “Performance evaluation of the deep learning based convolutional neural network approach for the recognition of chest X-ray images,” *Frontiers in oncology*, vol. 12, 2022.
- [57] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, “High-resolution image synthesis with latent diffusion models,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10684–10695, 2022.
- [58] E. J. Hu *et al.*, “LoRA: Low-rank adaptation of large language models,” *arXiv:2106.09685*, 2021.
- [59] V. Movčana *et al.*, “Organ-on-a-chip (OOC) image dataset for machine learning and tissue model evaluation,” *Data*, vol. 9, no. 2, p. 28, 2024.
- [60] J. Deng *et al.*, “ImageNet: A large-scale hierarchical image database,” in *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255, IEEE, 2009.
- [61] T.-Y. Lin *et al.*, “Microsoft COCO: Common objects in context,” in *Computer Vision—ECCV 2014. ECCV 2014. Lecture Notes in Computer Science*, vol. 8693, pp. 740–755, Springer, 2014.

Pateicības

Vēlos izteikt visdziļāko pateicību vairākiem cilvēkiem, kuru atbalsts un iedrošinājums bija ļoti būtisks šī darba pabeigšanā.

Pirmkārt, vēlos pateikties promocijas darba vadītājam Dr. Robertam Kadiķim par viņa vērtīgām norādēm, pieredzi un pacietību. Viņa atziņas un padomi palīdzēja man gan veidot manu zinātnisko karjeru kopumā, gan izstrādāt šo darbu.

Esmu bezgala pateicīgs savai ģimenei par atbalstu visa šī ceļa garumā: sievai Ilzei – par mīlestību, pacietību un kopā būšanu; meitām Laurai, Alisei, Adriānai, Emīlijai un Karolīnai – par to, ka viņas piepilda manu dzīvi ar prieku un iedvesmu; mātei Ninai – par ticību maniem panākumiem.

Īpašs paldies maniem kolēģiem Elektronikas un datorzinātņu institūtā (EDI), īpaši Dr. Modrim Greitānam un Dr. Kasparam Ozolam par labvēlīgas pētnieciskās vides radīšanu, kā arī Dr. Atim Elstam, Dr. Jānim Judvaitim un kolēģiem doktorantiem un grāda pretendentiem Krišjānim Nesenbergam, Didzim Lapsam un Anatolijam Zencovam par iespēju iesaistīties intelektuāli bagātinātās diskusijās.

Esmu ļoti pateicīgs visiem manu publikāciju līdzautoriem, ar kuriem man ir bijis prieks un gods sadarboties. Viņu ieguldījums ir bijis ļoti nozīmīgs manā darbā.

Vēlos pateikties arī saviem kolēģiem LU, īpaši profesoriem Jurim Borzovam, Zanei Bičevskai, Jānim Zuteram, Lailai Niedrītei un Guntim Arnicānam, par iespējām pasniegt LU, kā arī par vērtīgiem padomiem un palīdzību dažādos akadēmiskos un administratīvos jautājumos. Pateicos arī LU administratīvajam personālam, īpaši Ārijai Sproģei, Dacei Mileikai un mūžībā aizgājušajai Anitai Ermušai, par palīdzību dažādos organizatoriskos jautājumos. Kā arī esmu ļoti pateicīgs profesoram Jānim Zuteram un asociētajam profesoram Edgaram Celmam par to, ka viņi bija eksaminatori manā promocijas eksāmenā, profesoriem Uldim Straujumam, Guntim Arnicānam un Jurim Borzovam par vērtīgiem komentāriem šī darba agrākās versijas publiskās apspriešanas laikā, kā arī asociētajam profesoram Jevgēnijam Vihrovam, kurš kopīgoja ar mani sava promocijas darba \LaTeX veidni, tādējādi ietaupot man daudz laika un pūļu šī darba formatēšanā.

Izsaku sirsnīgu pateicību saviem Latvijas Universitātes studentiem, kuri ir devuši lielu ieguldījumu manā kā docētāja attīstībā, tādējādi padarot mani arī par labāku pētnieku.

Visbeidzot, vēlos pateikties maniem suņiem Pērlei, Karai (ak, mazā mīlā Kara!), Lesijai un Bellai, kā arī papagaiļiem Ronam, Lirai, Solo, Frodo un Tairai, kura diemžēl vairs nav ar mums, par viņu kopā būšanu

un prieku, ko viņi ienesuši manā dzīvē. Man viņi ir bijuši mierinājuma un sirdsmiera avots vissaspringtākajos brīžos.

Šis akadēmiskais ceļojums ir bijis izaicinošs, pateicīgs un brīžiem atkal izaicinošs, un esmu pateicīgs, ka man ir bijis tik atbalstošs ģimenes, kolēģu, studentu un draugu tīkls. Vēlreiz paldies jums visiem!