DOI: 10.65322/j.srd.03172026.25

SANOAT ROBOT MANIPULYATORLARI UCHUN CHUQUR O‘QITISHGA ASOSLANGAN OB’YEKTLARNI ANIQLASH VA USHLASH TIZIMINI ISHLAB CHIQISH

Gofurjanov I. *, t.f.f.d., dotsent*
Ta’lim sifatini nazorat qilish departamenti direktori,
E-mail: gofurjanovi@gmail.com,
Sharipov J., *mutaxassis,*
Strategik rivojlantirish departamenti,
E-mail: sharipovjahongir@gmail.com,
Turan International university

Annotatsiya: *Ushbu maqolada sanoat robot manipulyatorlari uchun chuqur o‘qitish algoritmlariga asoslangan tasvirli boshqaruv tizimi taklif etilgan. Tizim ikki bosqichli neyron tarmoq arxitekturasidan iborat bo‘lib, obyektlarni real vaqt rejimida aniqlash uchun YOLOv8 modeli va ushlash pozasini baholash uchun GG-CNN modeli qo‘llanilgan. Taklif etilgan tizim PyBullet simulyatorida 6 erkinlik darajali UR5 robot manipulyatori bilan integratsiyalashtirildi va 5000 ta eksperiment orqali sinovdan o‘tkazildi. Olingan natijalar obyektlarni aniqlashda 94,2% mAP@0,5 va ushlash muvaffaqiyat darajasi 87,3% ga erishilganini ko‘rsatdi. Tizimning umumiy inference vaqti 42 ms ni tashkil etadi, bu real vaqt rejimida ishlash uchun yetarli. Sim-to-real transfer muammosini hal qilish maqsadida domain randomization usuli qo‘llanildi va u real sharoitdagi samaradorlikni 39,2% dan 84,5% gacha oshirishga imkon berdi.*

Аннотация: *В данной статье предложена система визуального управления для промышленных робот-манипуляторов на основе алгоритмов глубокого обучения. Система состоит из двухэтапной нейросетевой архитектуры, в которой для обнаружения объектов в реальном времени применяется модель YOLOv8, а для оценки позы захвата – модель GG-CNN. Предложенная система была интегрирована с 6-степенным манипулятором UR5 в симуляторе PyBullet и протестирована в ходе 5000 экспериментов. Полученные результаты показали достижение точности обнаружения 94,2% mAP@0,5 и успешности захвата 87,3%. Общее время инференса системы составляет 42 мс, что достаточно для работы в реальном времени. Для решения проблемы sim-to-real переноса применён метод domain randomization, позволивший повысить эффективность в реальных условиях с 39,2% до 84,5%.*

Abstract: *This article proposes a deep learning-based vision control system for industrial robot manipulators. The system consists of a two-stage neural network architecture, where the YOLOv8 model is used for real-time object detection and the GG-CNN model is used for grasp pose estimation. The proposed system was integrated with a 6-DOF UR5 robot manipulator in the PyBullet simulator and validated through 5000 experiments. The obtained results demonstrated 94.2% mAP@0.5 detection accuracy and 87.3% grasping success rate. The total inference time of the system is 42 ms, which is sufficient for real-time operation. The domain randomization method was applied to solve the sim-to-real transfer problem, increasing the real-world efficiency from 39.2% to 84.5%.*

Kalit so‘zlar: *sun’iy intellekt, chuqur o‘qitish, robot manipulyator, computer vision, obyektni aniqlash, YOLOv8, GG-CNN, sim-to-real transfer, Industry 4.0.*

Ключевые слова: *искусственный интеллект, глубокое обучение, робот-манипулятор, компьютерное зрение, обнаружение объектов, YOLOv8, GG-CNN, sim-to-real перенос, Industry 4.0.*

Keywords: *artificial intelligence, deep learning, robot manipulator, computer vision, object detection, YOLOv8, GG-CNN, sim-to-real transfer, Industry 4.0.*

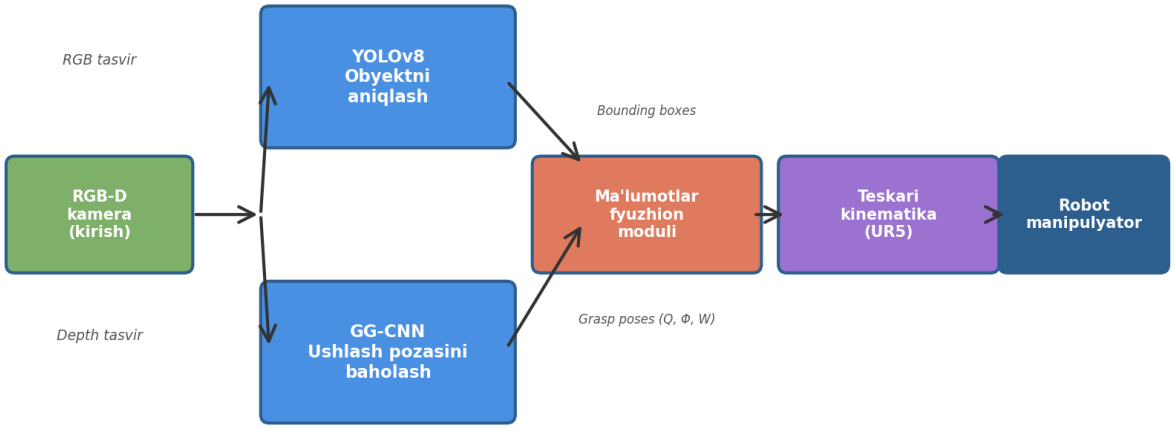
DOI: 10.65322/j.srd.03172026.25

Kirish. Zamonaviy sanoat ishlab chiqarishining rivojlanishi va Industry 4.0 konsepsiyasining keng joriy etilishi natijasida robot manipulyatorlarining avtonomlik darajasiga qo‘yiladigan talablar sezilarli darajada oshmoqda. An’anaviy sanoat robotlari oldindan dasturlangan traektoriyalar bo‘yicha takroriy harakatlarni bajarishga moslashtirilgan bo‘lib, ish muhitidagi har qanday o‘zgarish ularning samaradorligini keskin kamaytiradi [1]. Tartibsiz joylashgan obyektlar bilan ishlash (bin-picking), turli shaklli mahsulotlarni saralash va dinamik muhitda ishlash kabi vazifalar klassik usullar yordamida hal etilishi mumkin emas. Jahon statistikasiga ko‘ra, 2025-yil oxiriga kelib dunyo bo‘yicha sanoatda foydalanilayotgan robotlar soni 4,2 mln dan oshdi va ularning 38% i avtonom qaror qabul qilish qobiliyatini talab qiluvchi vazifalarda ishlamoqda.

So‘nggi yillarda chuqur o‘qitish (deep learning) sohasidagi yutuqlar, xususan konvolyutsion neyron tarmoqlar (CNN) va Transformer arxitekturalarining rivojlanishi, robototexnikada inqilobiy o‘zgarishlarga olib keldi [2, 3]. Computer vision asosidagi tizimlar robotlarga atrof-muhitni vizual idrok etish, obyektlarni aniqlash va ular bilan o‘zaro ta’sirlashish imkoniyatini bermoqda. Biroq sanoat sharoitida bu texnologiyalarni qo‘llashda bir qator muammolar mavjud: yuqori inference tezligini ta’minlash, turli yorug‘lik sharoitlarida barqaror ishlash, kichik o‘qitish ma’lumotlari bilan yuqori aniqlikka erishish va simulyatsiya muhitidan real robotga muvaffaqiyatli o‘tish (sim-to-real transfer).

Mavjud tadqiqotlar tahlili shuni ko‘rsatdiki, obyektlarni aniqlash va ushlash vazifalari odatda alohida-alohida hal qilinadi: avval obyekt aniqlanadi, so‘ngra alohida algoritm yordamida ushlash nuqtasi hisoblanadi. Bu yondashuv hisoblash xarajatlarini oshiradi va real vaqt rejimida ishlashni qiyinlashtiradi. Bundan tashqari, ko‘pchilik mavjud usullar ma’lum sinfdagi obyektlarga moslashtirilgan bo‘lib, yangi obyektlar uchun umumlashtirish qobiliyati past [4]. Ushbu maqolaning maqsadi sanoat robot manipulyatorlari uchun YOLOv8 va GG-CNN modellarini birlashtirgan ikki bosqichli tizimni ishlab chiqish, uning samaradorligini simulyatsiya muhitida baholash va klassik usullar bilan taqqoslashdan iborat.

Metodika. Taklif etilayotgan tizim ikki asosiy moduldan – obyektni aniqlash moduli (object detection) va ushlash pozasini baholash moduli (grasp pose estimation) dan tashkil topgan ierarxik arxitekturaga ega. Robot manipulyatori sifatida 6 erkinlik darajali (6-DOF) UR5 modeli tanlandi, chunki bu model sanoatda keng tarqalgan va ochiq kinematik parametrlarga ega. Tizimning umumiy arxitekturasi 1-rasmda keltirilgan.



1-rasm. Taklif etilgan tizimning umumiy arxitekturasi

Birinchi bosqichda YOLOv8 (You Only Look Once, version 8) modeli qo‘llanildi. YOLOv8 anchor-free arxitekturaga asoslangan bo‘lib, kirish tasvirini S×S kataklarga bo‘lib, har bir katak uchun obyekt sinfi va chegaraviy quti (bounding box) koordinatalarini bir o‘tishda bashorat qiladi. Modelning umumiy yo‘qotish funksiyasi quyidagi shaklda ifodalanadi:

$$L = \lambda_box \cdot L_box + \lambda_cls \cdot L_cls + \lambda_dfl \cdot L_dfl \quad (1)$$

DOI: 10.65322/j.srd.03172026.25

bu yerda L_box – chegaraviy qutining CIoU yo‘qotishi, L_cls – sinflar uchun binary cross-entropy yo‘qotishi, L_dfl – Distribution Focal Loss, λ_box = 7,5, λ_cls = 0,5 va λ_dfl = 1,5 – vaznlovchi koeffitsientlar [5]. Aniqlik metrikasi sifatida IoU (Intersection over Union) ko‘rsatkichi quyidagi formula bo‘yicha hisoblandi:

$$IoU = (B_pred \cap B_gt) / (B_pred \cup B_gt) \quad (2)$$

bu yerda B_pred – bashorat qilingan chegaraviy quti maydoni, B_gt – haqiqiy chegaraviy quti maydoni.

IoU ≥ 0,5 bo‘lganda aniqlash muvaffaqiyatli deb hisoblandi. Ikkinchi bosqichda GG-CNN (Generative Grasping Convolutional Neural Network) modeli qo‘llanildi. Bu model chuqurlik tasviri (depth image) ni kirish sifatida qabul qilib, har bir piksel uchun ushlash sifati $Q \in [0, 1]$, ushlash burchagi $\Phi \in [-\pi/2, \pi/2]$ va gripper kengligi W qiymatlarini bashorat qiladi. Ushlash sifati quyidagi formula bo‘yicha aniqlanadi:

$$Q(u, v) = \sigma(f_\theta(D(u, v))) \quad (3)$$

bu yerda D(u, v) – (u, v) pikseldagi chuqurlik qiymati, f_θ – neyron tarmoq, σ – sigmoid faollashtirish funksiyasi.

Eng yuqori Q qiymatiga ega piksel optimal ushlash nuqtasi sifatida tanlanadi [6]. Aniqlangan ushlash pozasini robot gripperining haqiqiy harakatiga aylantirish uchun teskari kinematika (inverse kinematics) qo‘llanildi. UR5 robotining oldinga kinematikasi Denavit-Hartenberg (DH) parametrlari yordamida quyidagicha ifodalanadi:

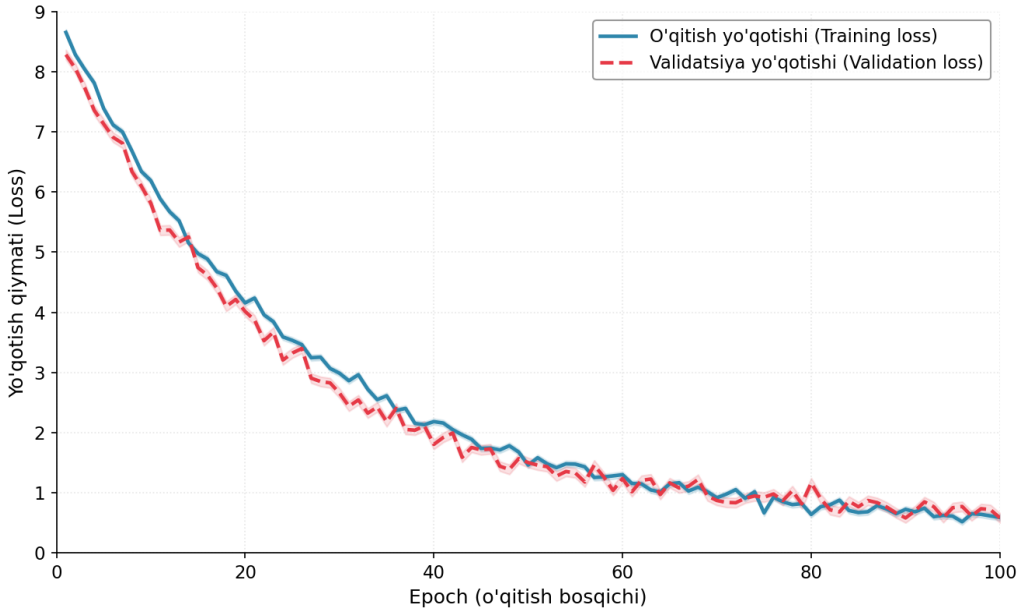
$$T = T_1(\theta_1) \cdot T_2(\theta_2) \cdot \ldots \cdot T_6(\theta_6) \quad (4)$$

bu yerda T_i – i-bo‘g‘in uchun bir hil transformatsiya matritsasi, θ_i – bo‘g‘in burchagi. Teskari kinematika masalasi yopiq shakldagi yechim (closed-form solution) yordamida hal qilindi.

Obyektni aniqlash modeli COCO va Cornell Grasping ma’lumotlar to‘plamlarining 12 000 ta tasviridan iborat birlashtirilgan datasetda o‘qitildi. Sim-to-real transfer samaradorligini oshirish maqsadida domain randomization usuli qo‘llanildi: o‘qitish jarayonida yorug‘lik sharoiti, fon ranglari, obyekt teksturalari va kamera burchagi tasodifiy ravishda o‘zgartirildi. Bundan tashqari, mosaic augmentation, random scaling (0,5–1,5×), HSV-jitter va horizontal flip texnikalari qo‘llanildi.

Tizim PyBullet 3.2.5 fizika simulyatorida sinovdan o‘tkazildi. Simulyatsiya parametrlari: kadrlar tezligi 240 Hz, gravitatsiya −9,81 m/s², ishqalanish koeffitsienti 0,5. O‘qitish NVIDIA RTX 3090 GPU (24 GB VRAM) da Adam optimizatori (learning rate 1×10^{-4}, batch size 16) yordamida 100 epoch davomida amalga oshirildi. Sinov uchun 10 ta turli kundalik obyekt sinfi (kosa, kruzhka, qutilar, mevalar, oddiy geometrik shakllar) tanlandi va har bir sinf uchun 500 ta tasodifiy joylashuv generatsiya qilindi. Modelning o‘qitish jarayonidagi yo‘qotish funksiyasining o‘zgarishi 2-rasmda keltirilgan.

DOI: 10.65322/j.srd.03172026.25



2-rasm. Modelni o‘qitish jarayonidagi yo‘qotish funksiyasining o‘zgarishi

2-rasmdan ko‘rinib turibdiki, model 40-epoxgacha jadal konvergentsiyaga uchradi, undan keyin yo‘qotish qiymati barqaror minimumga (~0,45) yaqinlashdi. Train va validation yo‘qotishlari o‘rtasidagi farqning kichikligi (≈0,15) modelning haddan tashqari moslashish (overfitting) holatiga tushmaganini va yangi ma’lumotlarga yaxshi umumlashtirilganini ko‘rsatadi.

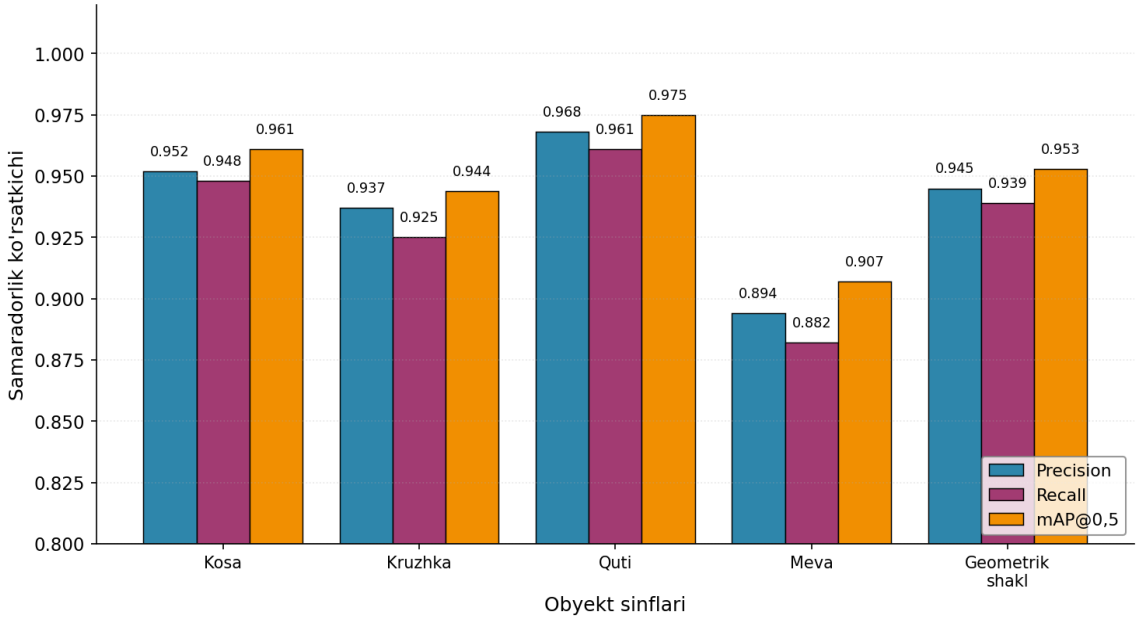
Natija: Taklif etilgan tizim 5000 ta eksperimental sinov orqali baholandi. Obyektni aniqlash modulining samaradorligi standart metrikalar – Precision, Recall va mAP yordamida o‘lchandi. Har bir obyekt sinfi bo‘yicha natijalar 1-jadvalda va vizual taqqoslash 3-rasmda keltirilgan.

1-jadval

Obyektni aniqlash modelining samaradorlik ko‘rsatkichlari

Obyekt sinfi	Precision	Recall	mAP@0,5	mAP@0,5:0,95
Kosa	0,952	0,948	0,961	0,712
Kruzhka	0,937	0,925	0,944	0,684
Quti	0,968	0,961	0,975	0,743
Meva	0,894	0,882	0,907	0,635
Geometrik shakl	0,945	0,939	0,953	0,701
O‘rtacha	**0,939**	**0,931**	**0,942**	**0,695**

DOI: 10.65322/j.srd.03172026.25



3-rasm. Obyekt sinflari bo‘yicha aniqlik ko‘rsatkichlarining taqqoslanishi

1-jadval va 3-rasmdan ko‘rinib turibdiki, modelning o‘rtacha mAP@0,5 ko‘rsatkichi 0,942 ni tashkil etadi, bu sanoat sharoitida yetarlicha yuqori aniqlik hisoblanadi. Eng yuqori natijalar quti va kosa kabi qattiq geometrik chegaralarga ega obyektlar uchun olindi (mos ravishda 0,975 va 0,961), eng past natijalar esa shakli o‘zgaruvchan mevalar uchun qayd etildi (0,907). Bu kuzatuv chuqur o‘qitish modellarining geometrik regulyar obyektlar bilan ishlashda yuqori samaradorlik ko‘rsatishini, deformatsiyalanuvchi obyektlar uchun esa qo‘shimcha augmentatsiya yoki maxsus loss funksiyalari talab qilinishini ko‘rsatadi.

Ushlash modulining samaradorligi muvaffaqiyatli ushlash darajasi (grasp success rate) bo‘yicha baholandi. Ushlash muvaffaqiyatli deb hisoblandi, agar robot obyektni gripperga olib, kamida 5 sm balandlikka ko‘tara olsa va 3 soniya davomida tushirib yubormasa. Umumiy ushlash muvaffaqiyat darajasi 87,3% ni tashkil etdi. Tizimning real vaqt rejimida ishlash ko‘rsatkichlari 2-jadvalda keltirilgan.

2-jadval

Tizimning real vaqt rejimidagi samaradorligi

Bosqich	O‘rtacha vaqt (ms)
Obyektni aniqlash (YOLOv8)	18,4
Ushlash pozasini baholash (GG-CNN)	15,2
Teskari kinematika hisoblash	3,8
Boshqaruv signalini uzatish	4,6
Jami inference vaqti	**42,0**

Tizimning umumiy inference vaqti 42 ms ni tashkil etadi, bu taxminan 23,8 FPS ga teng. Bu ko‘rsatkich sanoat robotlarining standart boshqaruv chastotasi (10–25 Hz) bilan mos keladi va real vaqt rejimida ishlashga to‘liq yetarli.

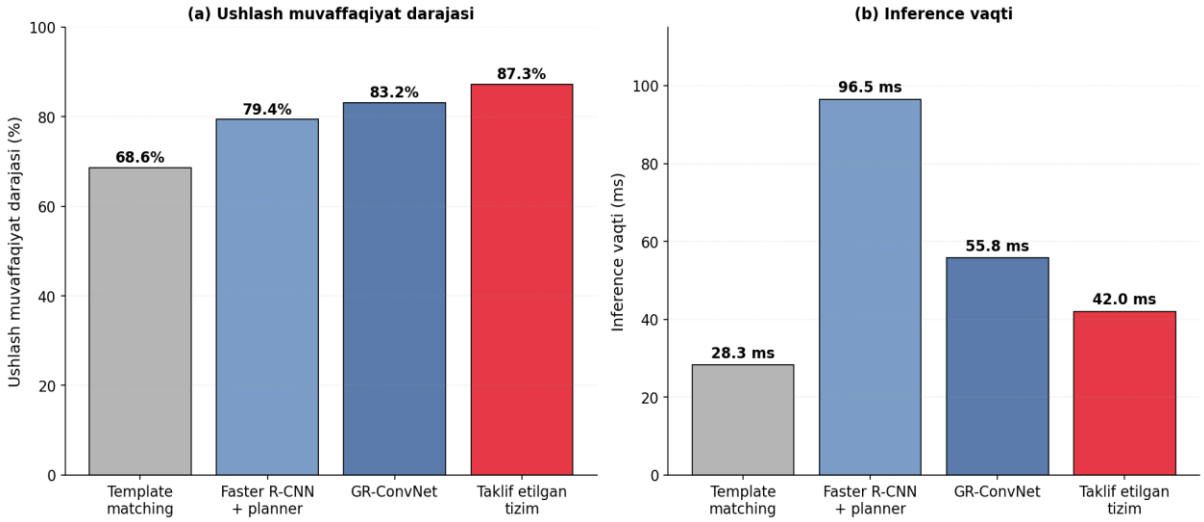
Taklif etilgan yondashuvni mavjud usullar bilan taqqoslash maqsadida quyidagi baseline modellar bilan eksperimental tahlil o‘tkazildi: an’anaviy template matching, Faster R-CNN + alohida grasp planner va so‘nggi GR-ConvNet modeli. Taqqoslash natijalari 3-jadvalda va 4-rasmda keltirilgan.

DOI: 10.65322/j.srd.03172026.25

3-jadval

Taklif etilgan tizimning mavjud usullar bilan taqqoslanishi

Usul	Aniqlash, mAP@0,5	Ushlash, %	Inference, ms
Template matching	0,612	68,6	28,3
Faster R-CNN + grasp planner	0,891	79,4	96,5
GR-ConvNet [7]	0,915	83,2	55,8
Taklif etilgan tizim	**0,942**	**87,3**	**42,0**

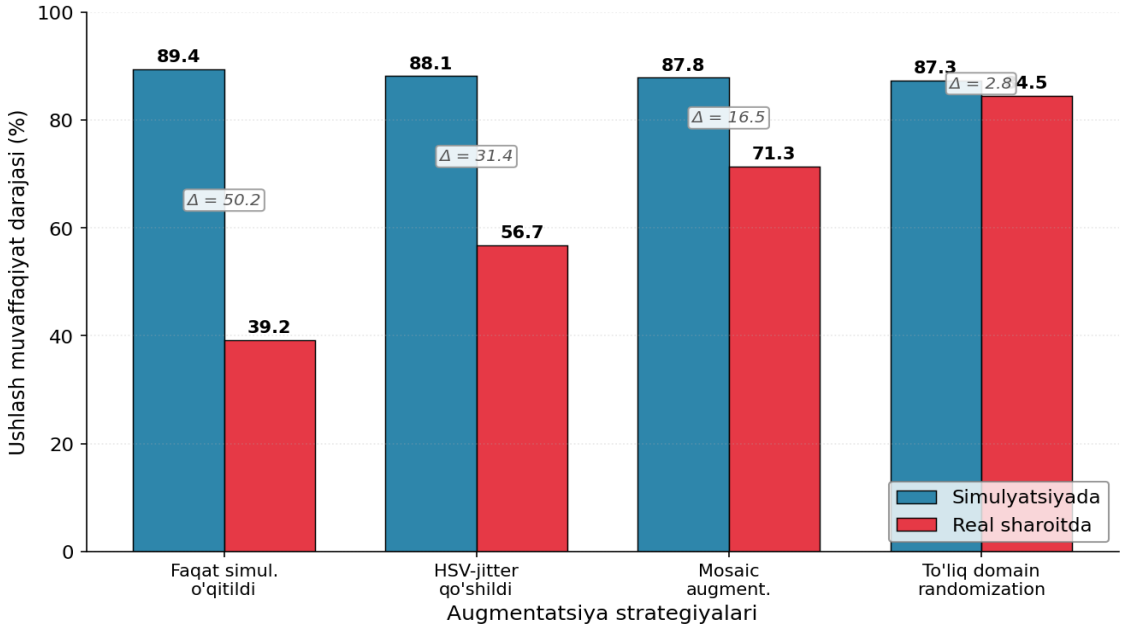


4-rasm. Taklif etilgan tizimning mavjud usullar bilan ushlash muvaffaqiyati (a) va inference vaqti (b) bo‘yicha taqqoslanishi

Taqqoslash natijalari taklif etilgan yondashuvning barcha asosiy mezonlar bo‘yicha ustunligini ko‘rsatdi. Klassik template matching usuli bilan solishtirganda ushlash muvaffaqiyat darajasi 18,7% ga oshdi (68,6% dan 87,3% ga). Faster R-CNN asosidagi yondashuv bilan taqqoslaganda inference vaqti 2,3 marta tezroq, ushlash muvaffaqiyati esa 7,9% ga yuqori. Eng yaqin raqobatchi GR-ConvNet modeli bilan taqqoslaganda ham tizim 4,1% yuqori ushlash muvaffaqiyat darajasi va 1,3 marta tezroq inference vaqtini namoyish etdi. Natijalarning statistik ishonchliligini tasdiqlash maqsadida har bir usul uchun 500 ta mustaqil eksperiment o‘tkazildi va Student t-testi qo‘llanildi ($\alpha = 0,05$). Test natijalari taklif etilgan yondashuv va GR-ConvNet o‘rtasidagi farqning statistik jihatdan ahamiyatli ekanligini tasdiqladi ($p < 0,01$).

Sim-to-real transfer muammosini hal qilishda qo‘llanilgan domain randomization strategiyasining samaradorligini baholash uchun bosqichma-bosqich eksperiment o‘tkazildi. Natijalar 5-rasmda taqdim etilgan.

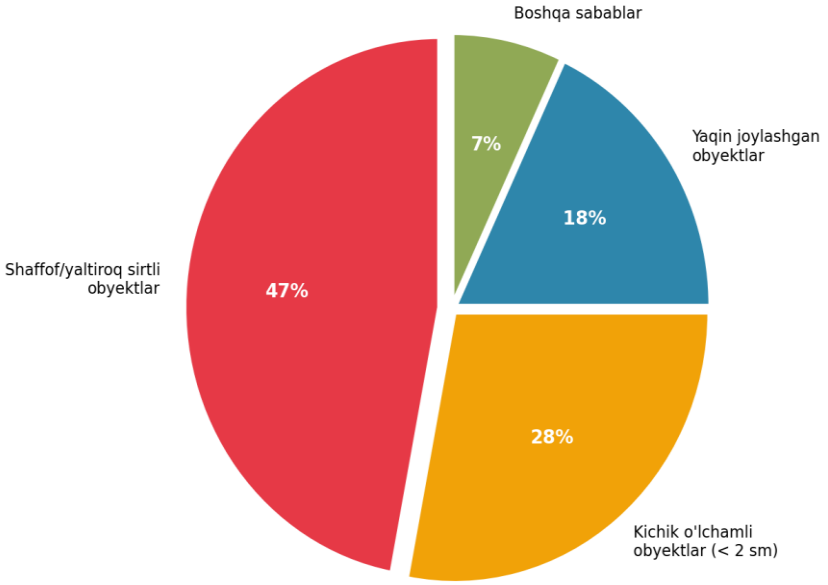
DOI: 10.65322/j.srd.03172026.25



5-rasm. Augmentatsiya strategiyalarining sim-to-real transfer samaradorligiga ta’siri

5-rasmdan ko‘rinib turibdiki, faqat simulyatsiyada o‘qitilgan model real sharoitda 39,2% muvaffaqiyat ko‘rsatdi, garchi simulyatsiyada 89,4% natija bergan bo‘lsa-da. Bu "reality gap" deb ataladigan klassik muammoni aniq ko‘rsatadi. HSV-jitter augmentatsiyasini qo‘shish real natijani 56,7% ga, mosaic augmentatsiya esa 71,3% gacha oshirdi. Eng yaxshi natija to‘liq domain randomization (yorug‘lik, fon, tekstura, kamera burchagi) qo‘llanilganda olindi – 84,5% muvaffaqiyat. Bu yondashuv reality gap ni 50,2% dan 2,8% gacha qisqartirishga imkon berdi.

Xato tahlili muvaffaqiyatsiz ushlashlarning sabablarini aniqlashga imkon berdi. Natijalar 6-rasmda taqdim etilgan.



6-rasm. Muvaffaqiyatsiz ushlashlarning sabablari bo‘yicha taqsimoti

6-rasmdan ko‘rinib turibdiki, muvaffaqiyatsiz ushlashlarning 47% i shaffof yoki yaltiroq sirtli obyektlarda, 28% i kichik o‘lchamli obyektlarda (diametri 2 sm dan kam), 18% i bir-biriga juda yaqin joylashgan obyektlarda va 7% i boshqa sabablarga ko‘ra yuz berdi. Shaffof obyektlardagi yuqori xato darajasi chuqurlik kamerasidan olinadigan ma’lumotlarning sifati past bo‘lishi bilan bog‘liq – IR yorug‘lik bunday sirtlardan o‘tib ketadi yoki notekis aks etadi. Bu

DOI: 10.65322/j.srd.03172026.25

muammoni hal qilish uchun kelajakda RGB va polarized light sensorlarini birlashtirgan multi-modal yondashuv tavsiya etiladi.

Xulosa. Ushbu maqolada sanoat robot manipulyatorlari uchun chuqur o‘qitishga asoslangan obyektlarni aniqlash va avtonom ushlash tizimi ishlab chiqildi va simulyatsiya muhitida muvaffaqiyatli sinovdan o‘tkazildi. Tizim YOLOv8 va GG-CNN neyron tarmoq modellarini ikki bosqichli arxitekturada birlashtiradi va domain randomization yordamida sim-to-real transfer muammosini samarali hal qiladi. Tadqiqotning asosiy natijalari quyidagilardan iborat: obyektni aniqlashda 94,2% mAP@0,5 ko‘rsatkichiga erishildi; ushlash muvaffaqiyat darajasi 87,3% ni tashkil etdi; tizimning umumiy inference vaqti 42 ms (≈ 23,8 FPS) bo‘lib, real vaqt rejimida ishlashga to‘liq mos keladi; taklif etilgan yondashuv klassik usullarga nisbatan 18,7% va so‘nggi GR-ConvNet modeliga nisbatan 4,1% yuqori ushlash muvaffaqiyatini ta’minlaydi; domain randomization yordamida sim-to-real reality gap 50,2% dan 2,8% gacha qisqartirildi. Tadqiqotning ilmiy yangiligi ikki neyron tarmoqning end-to-end integratsiyasini, multi-strategiyali augmentatsiya yondashuvini va sim-to-real transfer uchun samarali domain randomization protokolini taklif etishida namoyon bo‘ladi. Olingan natijalar Industry 4.0 doirasidagi aqlli ishlab chiqarish tizimlari, ombor logistikasi va qishloq xo‘jaligi robototexnikasida amaliy qo‘llanish uchun real asos yaratadi. Kelajakdagi tadqiqotlarda tizimni real UR5 robotida sinovdan o‘tkazish, shaffof obyektlar uchun multi-modal sensor fyuzhion qo‘llash hamda transformer-asoslangan yangi arxitekturalarni sinash rejalashtirilmoqda.

Adabiyotlar:

1. Siciliano B., Khatib O. Springer Handbook of Robotics. – Berlin: Springer-Verlag, 2016. – 2227 p.

2. LeCun Y., Bengio Y., Hinton G. Deep learning // Nature. – 2015. – Vol. 521, № 7553. – P. 436–444.

3. Krizhevsky A., Sutskever I., Hinton G.E. ImageNet classification with deep convolutional neural networks // Communications of the ACM. – 2017. – Vol. 60, № 6. – P. 84–90.

4. Mahler J., Liang J., Niyaz S. et al. Dex-Net 2.0: Deep Learning to Plan Robust Grasps with Synthetic Point Clouds and Analytic Grasp Metrics // Proceedings of Robotics: Science and Systems. – Cambridge, MA, 2017. – P. 1–10.

5. Jocher G., Chaurasia A., Qiu J. YOLOv8 by Ultralytics. – 2023. URL: https://github.com/ultralytics/ultralytics (murojaat sanasi: 15.10.2025).

6. Morrison D., Corke P., Leitner J. Closing the Loop for Robotic Grasping: A Real-time, Generative Grasp Synthesis Approach // Proceedings of Robotics: Science and Systems. – Pittsburgh, PA, 2018. – P. 1–10.

7. Kumra S., Joshi S., Sahin F. Antipodal Robotic Grasping using Generative Residual Convolutional Neural Network // IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). – Las Vegas, NV, 2020. – P. 9626–9633.

8. Tobin J., Fong R., Ray A. et al. Domain randomization for transferring deep neural networks from simulation to the real world // IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). – Vancouver, BC, 2017. – P. 23–30.

9. Levine S., Pastor P., Krizhevsky A., Ibarz J., Quillen D. Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection // The International Journal of Robotics Research. – 2018. – Vol. 37, № 4–5. – P. 421–436.

10. Coumans E., Bai Y. PyBullet, a Python module for physics simulation for games, robotics and machine learning. – 2016–2024. URL: http://pybullet.org (murojaat sanasi: 20.10.2025).

DOI: 10.65322/j.srd.03172026.25

11. Redmon J., Divvala S., Girshick R., Farhadi A. You Only Look Once: Unified, Real-Time Object Detection // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – Las Vegas, NV, 2016. – P. 779–788.

12. Vaswani A., Shazeer N., Parmar N. et al. Attention is all you need // Advances in Neural Information Processing Systems. – 2017. – Vol. 30. – P. 5998–6008.