



ELTE
EÖTVÖS LORÁND
UNIVERSITY

Akadémiai-Ipari Együtműködési Központ
Informatikai Kar
Eötvös Loránd Tudományegyetem

Adatalapú Robotika Transzfer és Curriculum Tanulással

DOKTORI ÉRTEKEZÉS TÉZISEI

HORVÁTH Dániel

Témavezető: Dr. ISTENES Zoltán, PhD

Ipari Témavezető: Dr. ERDŐS Ferenc Gábor, PhD

Campus France Internship Témavezető: Prof. MOUTARDE Fabien, PhD

Informatika Doktori Iskola

Iskolavezető: Prof. HORVÁTH Zoltán, PhD

Információs Rendszerek Doktori Program

Programvezető: Prof. BENCZÚR András, PhD

Budapest, 2024

Kivonat

A mély tanulás forradalma alapjaiban változtatta meg számos tudományterületet, beleértve a robotikát is. Azonban, ahhoz, hogy a mélytanulási algoritmusok lehetőségeit kiaknázzuk és adaptív, intelligens robotokat hozzunk létre, bizonyos kihívásokat le kell küzdeni. A nehézség elsősorban az érzékelési és mozgási képességek elsajátításában rejlik mintsem a felnőtt szintű intelligencia elérésében (Moravec-paradoxon). A legfőbb kihívások a transzferabilitás és az univerzalitás. A dolgozatban ezen kihívásokra koncentrálva, a transzferés és a curriculum tanulás (CL) segítségével kívánok megoldásokat adni. Fő eredményeim a következők.

Strukturálatlan környezetben működő robotoknak képeseknek kell lenniük észlelni és értelmezniük a környezetüket. A mélytanulási modellek alkalmazásának egyik fő akadálya a feladat-specifikus címkézett adatok hiánya. A valóság és a szimuláció közötti különbség áthidalása érdekében kifejlesztettem egy domén randomizáción alapuló sim2real tudástranszfer módszertant objektum detektáláshoz (S2R-ObjDet), amely lehetővé teszi címkézett szintetikus adatok automatikus generálását. Továbbá, javaslom a generalizált zavar-mátrixot (GCM) ami a klasszikus precizitás-szenzitivitás alapú metrikák hiányosságaira kínál megoldást. Ezenkívül készítettem egy nyilvános és annotált, valós ipari tárgyakkal álló adatbázist (InO-10-190).

Objektumok mozgásához nem csak a pozíciójuk, hanem az orientációjuk ismerete is szükséges. Ennek érdekében, két, gépi látáson alapuló, több objektum egyidejű megfogási helyzetének becslésére alkalmas modellt javaslok – a valós idejű MOGPE-RT és a nagy pontosságú MOGPE-HP modelleket. Továbbá az S2R-PosEst módszertant mely az S2R-ObjDet metódus kiterjesztése orientáció becslésre. Ez a keretrendszer ipari eszközt biztosít a gyors adatgeneráláshoz és modelltréninghez, miközben minimális valós adatot igényel.

A megerősítéses tanítás (RL) – az emberi tanulást mintául véve – univerzális megoldást kíván nyújtani különféle problémákra. Ugyanakkor a robotika jelentős kihívásokat állít. A robotágensek hatékony tanításának érdekében javaslom a kiemelt tapasztalati puffer (HiER), adatkiaknázó CL módszeremet, mely jelentősen javítja a state-of-the-art módszerek teljesítményét. Ennek továbbfejlesztésére, bevezetem a HiER+ módszert, amelyben a HiER egy tetszőleges adatgyűjtési CL módszerrel egészül ki, például az általam javasolt easy2hard kezdeti állapot entrópia módszerrel (E2H-ISE).

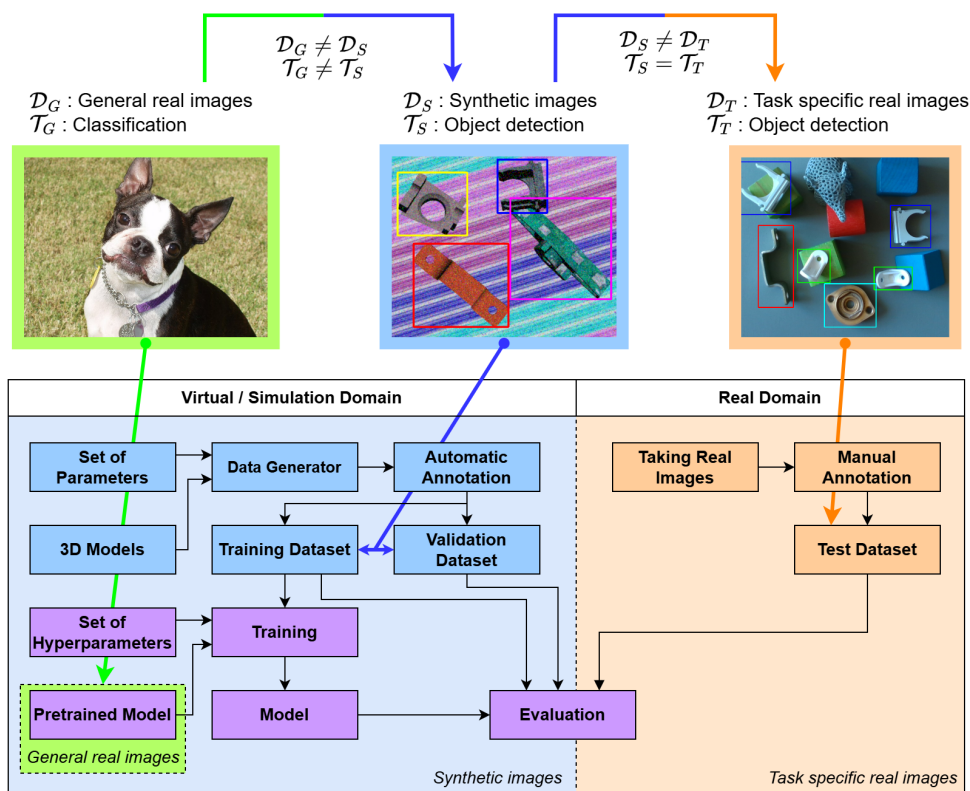
Bár a disszertációban bemutatott eredmények a saját munkám, a továbbiakban stilisztikai okokból a többes szám első személyt fogom használni. Az implementációk, kvalitatív eredmények, videó prezentációk és további anyagok elérhetők a disszertáció weboldalán: www.danielhorvath.eu/thesis.

Bevezetés

A mélytanulás (DL) a mesterséges intelligencia (AI) forradalom zászlóshajója, mely gyökeresen alakított át számos tudományterületet, beleértve a robotikát is. Azonban számos kihívást meg kell oldani ahhoz, hogy ki lehessen használni a DL algoritmusok teljes potenciálját, és valóban adaptív, intelligens robotokat lehessen létrehozni. Ez a dolgozat a robotika adat-alapú megközelítésének néhány kihívására igyekszik megoldást nyújtani, különös tekintettel a transzfer és a curriculum tanulásra. Az oldalszám korlát miatt, a részletesebb bevezetőt a disszertáció első fejezetében találja az olvasó.

Sim2real tudástranszfer objektumdetektáláshoz

Strukturálatlan környezetben működő robotoknak tudniuk kell észlelni és értelmezni a környezetüket. A mélytanulási modellek alkalmazásának egyik fő akadály a feladat-specifikus címkézett adatok hiánya. Így az első kutatási kérdésünk a következő: **Hogyan vihető át tudás a szimulációból a valóságba objektumdetektálás esetében?** Az első kutatási kérdéssel kapcsolatos tézisek a következők:



1. ábra. **Fent.** A tudástranszfer folyamatábrája. **Lent.** Az adatgenerálás, tanítás és kiértékelés folyamatábrája. Az ábrán lévő Boston terrier kép az ImageNet adatbázisból származik. [9].

I. Tézis: A sim2real domén randomizációs módszerünkkel (S2R-ObjDet) olyan szintetikus képeket lehet generálni, melyek segítségével az objektumdetektáló modellek az objektumok általános reprezentációját tanulják meg, ezzel áthidalva szimuláció és a valós környezet közötti különbséget.

Az S2R-ObjDet domén randomizáción alapuló objektumdetektáláshoz alkalmazható sim2real szintetikus adatgenerálási módszerünk során a megadott objektumok 3D modelljeit betöltjük a szimulátorba, továbbá minden objektum véletlenszerű textúrát vagy monokróm színt kap. Az objektumok száma és típusa is véletlenszerűen kerül kiválasztásra. A gravitációs erő szimulálásával az objektumokat egy síkra ejtjük, ahol az egyik stabil egyensúlyi helyzetükbe állnak be. A kamera külső és belső paramétereit is véletlenszerűen állítjuk be ügyelve arra, hogy az adott objektumok a látómezőben legyenek. A kép elkészítése után egy utófeldolgozási módszert alkalmazunk, amely többszínű só-bors zajt, Gauss-elmosást és opcionálisan téglalap, kör vagy vonal kivágásokat tartalmaz. A kép annotációját egy automatikus algoritmus végzi az objektumok összes pontja alapján, nem pedig a hagyományos, tengelyekhez igazított, 8 pontos befoglaló dobozok segítségével. Ez a folyamat addig ismétlődik, amíg a szükséges mennyiségű kép el nem készül a tanításhoz. Az S2R-ObjDet metódus segítségével a szimuláció és a valós környezet közötti különbség megfelelő szintre csökkenthető. A nyilvánosan elérhető, kézzel annotált InO-10-190 adathalmazunkon – amely 190 valós képet tartalmaz, összesen 920 objektumról, 10 osztályba sorolva – 86,32%-os és 97,38%-os mAP_{50} eredményeket értünk el zero-shot és one-shot transzfer esetén. Az osztályok között egymástól különböző és kifejezetten hasonló munkadarabok egyaránt szerepelnek annak érdekében, hogy az eltérő osztályok felismerése, illetve a hasonló objektumok megkülönböztetése egyaránt tesztelhető legyen. Megoldásunk megfelel az ipari igényeknek is, mivel egy kép generálása kevesebb, mint 0,5 másodpercet igényel, ezáltal gyors tanítási folyamatot biztosít. A módszertan folyamatábrája az 1. ábrán látható. Ez a tézis a következő publikációhoz kapcsolódik: [1].

II. Tézis: Az objektumdetektáláshoz tartozó klasszifikációs hibák, hamis pozitív és hamis negatív esetek – melyeket a hagyományos metrikák nem fednek le – hatékonyan mérhetők és kiértékelhetők az általunk javasolt generalizált zavarmátrix (GCM) segítségével.

Az általunk javasolt generalizált zavarmátrix (GCM) – amely a 2. ábrán látható – a klasszikus zavarmátrix objektumdetektálásra való adaptációja. A GCM megoldást kínál a hagyományos precizitás-szenzitivitás alapú mAP és F_1 metrikák korlátaira. Segítségével a klasszifikációs hibák, illetve a hamis pozitív és hamis negatív esetekből származó hibák hatékonyan mérhetők és kiértékelhetők. A hagyományos zavarmátrixhoz képest $\mathbf{D} \in \mathbb{N}^{C \times C}$, ahol $C \in \mathbb{N}$ az osztályok száma, a GCM-ben $\mathbf{D}^{gen} \in \mathbb{N}^{C+1 \times C+1}$ egy extra sor és egy extra oszlop kerül hozzáadásra a hamis pozitív és hamis negatív esetekhez. A helyes felismerések az átlóban találhatók, $D_{i,i}^{gen}$, ugyanúgy, mint a standard zavarmátrixban. $D_{C+1,C+1}^{gen} \doteq 0$. Ez a tézis a következő publikációhoz kapcsolódik: [1].

Az implementációnk, a kvantitatív és a kvalitatív kiértékelésünk, egy videóprezentáció és további anyagok a projekt weboldalán találhatóak: www.danielhorvath.eu/sim2real.

L-bracket	42	0	0	0	0	0	0	0	0	0	36
U-bracket	0	74	0	0	0	0	0	5	0	2	16
angle_bracket	0	0	89	1	0	0	0	0	1	0	15
seat	0	0	0	52	0	1	0	8	0	0	49
pipe_clamp	3	0	0	2	67	0	0	0	0	0	7
handle	0	0	0	0	0	71	6	0	0	0	12
bonnet	0	0	0	0	0	0	16	62	0	0	6
body	0	0	0	0	0	0	2	82	1	0	3
ball	0	0	0	0	0	0	0	0	83	0	4
cable_shoe	0	0	0	0	0	0	0	0	0	91	12
[None]	12	0	0	0	0	0	4	5	1	1	0
	L-bracket	U-bracket	angle_bracket	seat	pipe_clamp	handle	bonnet	body	ball	cable_shoe	[None]

2. ábra. Generalizált zavarmátrix (GCM).

Sim2real megfogási helyzetbecslés

Az előző részben bemutattuk a sim2real domén randomizációs módszerünket objektumdetektálás esetére. Azonban az objektumok mozgatásához nem csak a pozíciójuk, hanem az orientációjuk ismerete is szükséges. Így a második kutatási kérdésünk a következő: **Hogyan terjeszthető ki az S2R-ObjDet módszerünk több objektum megfogási helyzetének becslésére?** A második kutatási kérdéssel kapcsolatos tézisek a következők:

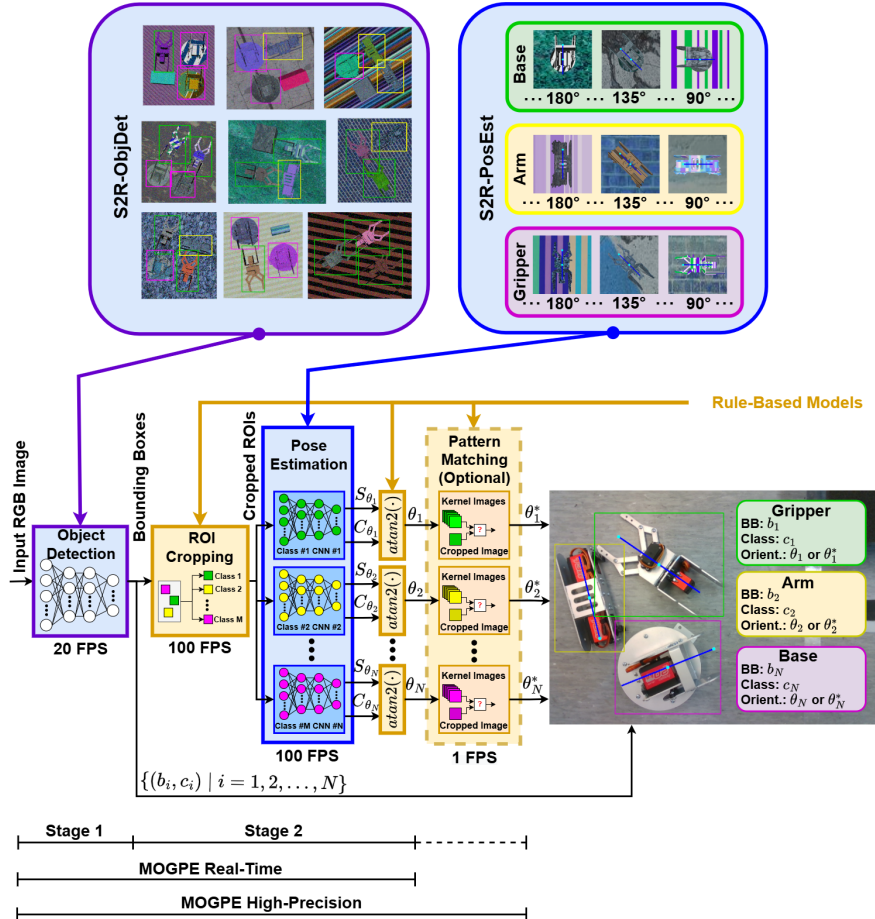
III. Tézis: A kétlépcsős, objektumok megfogási helyzetének egyidejű becslésére alkalmas modelljeink –a valós idejű MOGPE-RT és a nagy pontosságú MOGPE-HP – lehetővé teszik több objektum megfogási helyzetének egyidejű becslésének moduláris betanítását, az objektumdetektálás és osztályspecifikus orientációbecslés szekvenciális fázisainak alkalmazásával.

Két, gépi látás alapú, objektumok megfogási helyzetének egyidejű becslésére alkalmas modellt javaslunk – a valós idejű MOGPE-RT és a nagy pontosságú MOGPE-HP modelleket – amelyeket a 3. ábra mutat. Mindkét modell két fő komponensre épül: egy objektumdetektáló modellre és egy orientációbecslő modellre. Az objektumdetektáló modell kimenete $\mathbf{y} = \{(\mathbf{b}_i, c_i^{class}, p_i^{con}) \mid i = 1, 2, \dots, N\}$, ahol $\mathbf{b}_i = [x_i, y_i, w_i, h_i] \in [0, 1]^4$ az i -edik detekció tengelyhez igazított befoglaló dobozát jelöli, $c_i^{class} \in \mathbb{N}$ az i -edik detekció osztálya, $p_i^{con} \in [0, 1]$ az i -edik detekció bizalmi szintje, $N \in \mathbb{N}$ pedig a detektált objektumok száma. Azok a detekciók kiszűrésre kerülnek, amelyeknél $p_i^{con} < \tau_{con}$, ahol $\tau_{con} \in [0, 1]$ a bizalmi küszöbérték. A ROI kivágó modul kimetszi a képből az adott objektumokat, illetve átméretezi őket a kívánt méretre. Az osztályspecifikus orientációbecslő modellek kiszámítják a $\sin(\theta_i)$ és $\cos(\theta_i)$ értékeket minden objektumra, ahol $\theta_i \in [-\pi, \pi]$ az orientációs szög. Ezután az atan2 függvénnyel kiszámítjuk a θ_i szöveget, amely a MOGPE-RT modell kimenete. A MOGPE-HP modell egy további lokális mintázat-illesztő algoritmus alkalmazásával egy pontosabb $\theta^ \in [-\pi, \pi]$ szöveget számít ki. Ez a tézis a következő publikációhoz kapcsolódik: [4].*

IV. Tézis: Az S2R-PosEst módszerünk gyors szintetikus adatgenerálást tesz lehetővé az osztályspecifikus orientációbecslő modellek számára, hatékonyan áthidalva a valóság és a szimuláció közötti különbséget.

A javasolt S2R-PosEst módszer, egy sim2real domén randomizációs metódus orientációbecsléshez, amely az S2R-ObjDet módszerünkre épül. Az objektum 3D modelljét a szimulátorban helyezzük, majd a z-tengely körül forgatjuk – merőlegesen arra a síkra, amelyen az objektum található – miközben véletlenszerű textúrákat adunk a síknak és az objektumnak egyaránt. Összesen $n_{rot} = \lfloor \frac{2 \cdot \pi}{\beta_{res}} \rfloor$ forgatás történik, ahol $n_{rot} \in \mathbb{N}$ a forgatások száma, és $\beta_{res} \in \mathbb{R}$ a radiánban mért felbontás. Minden forgatáshoz egy kép készül, és a címke automatikusan generálódik hozzá. Az adatgenerálás képenként 0,25–0,5 másodpercet igényel, így ipari felhasználásra is alkalmas. Ez a tézis a következő publikációhoz kapcsolódik: [4].

Az implementációnk, a kvantitatív és a kvalitatív kiértékelésünk, egy videóprezentáció és további anyagok a projekt weboldalán találhatóak: www.danielhorvath.eu/mogpe.



3. ábra. **Fent.** Az S2N-ObjDet és S2N-PosEst módszereink illusztrációja. **Lent.** A több objektum megfogási helyzetének becslésére alkalmas modelljeink (MOGPE) folyamatábrája.

HiER: Kiemelt tapasztalati puffer

Az előzőekben a hangsúly a sim2real tudástranszferen volt felügyelt tanulási feladatok esetében. Mindazonáltal az adaptív robotika célja nem csupán a transzferabilitás, hanem az univerzalitás is, mely fogalmak között erős kapcsolat van, ugyanis az univerzális megoldások – definíció szerint – könnyen transzferálhatóak. Az univerzális megoldások felé vezető úton az egyik fontos építőelem a megerősítéses tanulás (RL) lehet. Hasonlóan az emberekhez, az RL algoritmusok is kísérletezés (*trial and error*) során, a környezettel való interakciókon keresztül tanulnak. A felügyelt tanulással összehasonlítva, az RL különösen hasznos a nagy ügyességet igénylő robotikai feladatoknál. Ugyanakkor a robotika területén jelentős kihívásokkal kell szembenézni, mivel az állapot- és akcióterek folytonosak, a jutalomfüggvény pedig többnyire ritka. Továbbá sok esetben az ágens nem fér hozzá semmilyen demonstrációhoz. Ezek alapján a harmadik kutatási kérdésünk: **Hogyan lehet továbbfejleszteni a state-of-the-art RL algoritmusok tanítását curriculum tanulás segítségével?** A harmadik kutatási kérdéssel kapcsolatos tézisek a következők:

V. Tézis: A kiemelt tapasztalati puffer (HiER) módszerünk jelentősen javítja a state-of-the-art megerősítéses tanuláson alapuló ágensek tanítását a legfontosabb tapasztalatok szeparált tárolásával és újrajátszásával.

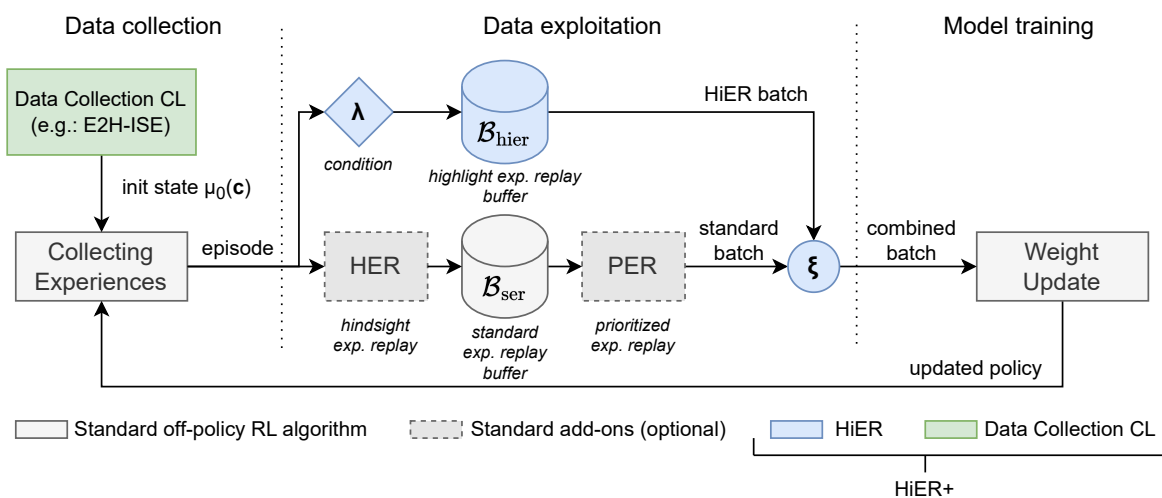
A kiemelt tapasztalati puffer (HiER) módszerünket az emberi tanulás inspirálta. Egy másodlagos tapasztalati visszajátszó puffert hoztunk létre, amely a legrelevánsabb tapasztalásokat tárolja. A tanítás során a tapasztalásokat mind a standard tapasztalati visszajátszó pufferből, mind a kiemelt tapasztalati visszajátszó pufferből mintavételezzük. Ez a módszer bármely off-policy RL ügynökhöz hozzáadható, és alkalmazható akár a hindsight experience replay (HER), akár a prioritized experience replay (PER) technikákkal, vagy nélkülük. A HiER a 4. ábrán látható és az 1. algoritmusban^a van részletezve. A HiER az adatkínázó vagy implicit curriculum tanulási módszerek osztályába tartozik. Amennyiben csak pozitív tapasztalatokat tárolunk a pufferében, automatikus demonstráció-generátorként is tekinthetünk rá. A HiER jelentősen javítja az RL referenciamódszerek teljesítményét, sztochasztikus dominanciát mutatva a state-of-the-art algoritmusokkal szemben, amelyet három robotikai benchmark nyolc feladatán validáltunk. Ez a tézis a következő publikációhoz kapcsolódik: [2].

^aA terjedelmi korlát miatt lásd a disszertációban.

VI. Tézis: A HiER+ módszerünk tovább javítja a kiemelt tapasztalati puffer (HiER) módszerünk teljesítményét azáltal, hogy növeli a pozitív tapasztalatok elérhetőségét – amit a feladat nehézségének szabályozásával érünk el – különösen a tanítás korai szakaszaiban.

A HiER+ módszerünk, a HiER kiegészítése egy tetszőleges adatgyűjtési (hagyományos) curriculum tanulási módszerrel. A HiER+ áttekintése a 4. ábrán látható és a 2. algoritmusban^a van részletezve. Továbbá az adatgyűjtési CL módszer példaként javasoljuk az E2H-ISE algoritmust, mely egy univerzális, könnyen megvalósítható easy2hard adatgyűjtési CL módszer, amely a feladat minimális ismeretét igényli. Az E2H-ISE algoritmus szabályozza a kezdeti kiinduló állapot és cél entrópia eloszlását $\mathcal{H}(\mu_0)$, mely közvetetten befolyásolja a feladat nehézségét. Kísérleti eredményeink azt mutatják, hogy a HiER+ tovább javítja a HiER teljesítményét. Ezenkívül a HiER+ sztochasztikus dominanciát mutat a HiER-rel szemben a Panda-Gym benchmark három robotikai feladatának eredményei alapján. Ez a tézis a következő publikációhoz kapcsolódik: [2].

^aA terjedelmi korlát miatt lásd a disszertációban.



4. ábra. HiER és HiER+ áttekintése.

Az implementációnk, a kvantitatív és a kvalitatív kiértékelésünk, egy videoprezentáció és további anyagok a projekt weboldalán találhatóak: www.danielhorvath.eu/hier

Publikációk

A szerző folyóiratcikkei

- [1] **D. Horváth**, G. Erdős, Z. Istenes, T. Horváth és S. Földi, “Object Detection Using Sim2Real Domain Randomization for Robotic Applications”, *IEEE Transactions on*

Robotics, 39. évf., 2. sz., 1225–1243. old., 2023. ápr., ISSN: 1941-0468. DOI: 10.1109/TR0.2022.3207619.

- [2] **D. Horváth**, J. Bujalance Martín, F. Gábor Erdős, Z. Istenes és F. Moutarde, “HiER: Highlight Experience Replay for Boosting Off-Policy Reinforcement Learning Agents”, *IEEE Access*, 12. évf., 100102–100119. old., 2024. júl., ISSN: 2169-3536. DOI: 10.1109/ACCESS.2024.3427012.
- [3] G. Erdős, K. Abai, R. Beregi és tsai., “Enabling Technologies for Autonomous Robotic Systems in Manufacturing”, *Transactions of Nanjing University of Aeronautics and Astronautics*, 41. évf., 4. sz., 403–431. old., 2024. aug., ISSN: 1005-1120. DOI: 10.16356/j.1005-1120.2024.04.001.

A szerző konferenciaközleményei

- [4] **D. Horváth**, K. Bocsi, G. Erdős és Z. Istenes, “Sim2Real Grasp Pose Estimation for Adaptive Robotic Applications”, *the 22nd IFAC World Congress*, IFAC-PapersOnLine sor., 56. köt., 2023, 5233–5239. old. DOI: 10.1016/j.ifacol.2023.10.121.
- [5] G. Erdős, **D. Horváth** és G. Horváth, “Visual Servo Guided Cyber-Physical Robotic Assembly Cell”, *the 17th IFAC Symposium on Information Control Problems in Manufacturing (INCOM)*, IFAC-PapersOnLine sor., 54. köt., 2021. jan., 595–600. old. DOI: 10.1016/j.ifacol.2021.08.068.
- [6] M. Hajós és **D. Horváth**, “Robotos Pakolási Feladat Megoldása Környezetérzékelés Segítségével”, *Nemzetközi Gépészeti Konferencia (OGÉT)*, 2020. ápr., 305–308. old. cím: <https://ojs.emt.ro/oget/article/view/156>.
- [7] Z. Kemény, R. Beregi, J. Nacsa, C. Kardos és **D. Horváth**, “Human–Robot Collaboration in the MTA SZTAKI Learning Factory Facility at Győr”, *the 8th CIRP Sponsored Conference on Learning Factories (CLF)*, Procedia Manufacturing sor., 23. köt., 2018. jan., 105–110. old. DOI: 10.1016/j.promfg.2018.04.001.
- [8] Z. Kemény, R. Beregi, J. Nacsa, C. Kardos és **D. Horváth**, “Example of a Problem-to-Course Life Cycle in Layout and Process Planning at the MTA SZTAKI Learning Factories”, *the 9th Conference on Learning Factories (CLF)*, Procedia Manufacturing sor., 31. köt., 2019. jan., 206–212. old. DOI: 10.1016/j.promfg.2019.03.033.

A disszertáció kiemelt hivatkozásai

- [9] J. Deng, W. Dong, R. Socher és tsai., “ImageNet: A Large-Scale Hierarchical Image Database”, *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009. jún., 248–255. old. DOI: 10.1109/CVPR.2009.5206848.
- [10] I. Goodfellow, Y. Bengio és A. Courville, *Deep Learning*. The MIT Press, 2016, ISBN: 9780262035613.

- [11] R. S. Sutton és A. G. Barto, *Reinforcement Learning: An Introduction*. A Bradford Book, 2018, ISBN: 9780262039246.
- [12] A. I. Károly, P. Galambos, J. Kuti és I. J. Rudas, “Deep Learning in Robotics: Survey on Model Structures and Training Strategies”, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 51. évf., 1. sz., 266–279. old., 2021. jan., ISSN: 2168-2232. DOI: 10.1109/TSMC.2020.3018325.
- [13] F. Zhuang, Z. Qi, K. Duan és tsai., “A Comprehensive Survey on Transfer Learning”, *Proceedings of the IEEE*, 109. évf., 1. sz., 43–76. old., 2021. jan., ISSN: 1558-2256. DOI: 10.1109/JPROC.2020.3004555.
- [14] T. Haarnoja, A. Zhou, P. Abbeel és S. Levine, “Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor”, *arXiv*, 2018. aug. DOI: 10.48550/arXiv.1801.01290.
- [15] X. Wang, Y. Chen és W. Zhu, “A Survey on Curriculum Learning”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44. évf., 9. sz., 4555–4576. old., 2022. szept., ISSN: 1939-3539. DOI: 10.1109/TPAMI.2021.3069908.
- [16] R. Portelas, C. Colas, L. Weng, K. Hofmann és P.-Y. Oudeyer, “Automatic Curriculum Learning For Deep RL: A Short Survey”, *arXiv*, 2020. máj. DOI: 10.48550/arXiv.2003.04664.
- [17] Y. Bengio, J. Louradour, R. Collobert és J. Weston, “Curriculum Learning”, *Proceedings of the 26th Annual International Conference on Machine Learning (ICML)*, Association for Computing Machinery, 2009. jún., 41–48. old., ISBN: 9781605585161. DOI: 10.1145/1553374.1553380.
- [18] J. Tobin, R. Fong, A. Ray és tsai., “Domain Randomization for Transferring Deep Neural Networks from Simulation to the Real World”, *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017. szept., 23–30. old. DOI: 10.1109/IROS.2017.8202133.
- [19] A. Barisic, F. Petric és S. Bogdan, “Sim2Air - Synthetic Aerial Dataset for UAV Monitoring”, *IEEE Robotics and Automation Letters*, 7. évf., 2. sz., 3757–3764. old., 2022. ápr., ISSN: 2377-3766. DOI: 10.1109/LRA.2022.3147337.
- [20] T. Schaul, J. Quan, I. Antonoglou és D. Silver, “Prioritized Experience Replay”, *arXiv*, 2016. febr. DOI: 10.48550/arXiv.1511.05952.
- [21] M. Andrychowicz, F. Wolski, A. Ray és tsai., “Hindsight Experience Replay”, *Advances in Neural Information Processing Systems*, 30. köt., Curran Associates, Inc., 2017. júl. DOI: 10.48550/arXiv.1707.01495.
- [22] C. Florensa, D. Held, M. Wulfmeier, M. Zhang és P. Abbeel, “Reverse Curriculum Generation for Reinforcement Learning”, *arXiv*, 2018. júl. DOI: 10.48550/arXiv.1707.05300.
- [23] J. Tremblay, A. Prakash, D. Acuna és tsai., “Training Deep Networks with Synthetic Data: Bridging the Reality Gap by Domain Randomization”, *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2018. jún., 1082–10828. old. DOI: 10.1109/CVPRW.2018.00143.