ELTE
EÖTVÖS LORÁND
UNIVERSITY

CoLocation Center for Academic and Industrial Cooperation
Faculty of Informatics
Eötvös Loránd University

Advancing Data-Driven Robotics with Transfer and Curriculum Learning

THESIS BOOKLET

Dániel Horváth

Supervisor: Dr. Zoltán Istenes, PhD

Industrial Supervisor: Dr. Ferenc Gábor Erdős, PhD

Campus France Internship Supervisor: Prof. Fabien Moutarde, PhD

Doctoral School of Informatics
Head: Prof. Zoltán Horváth, PhD

Doctoral Program of Information Systems
Head: Prof. András Benczúr, PhD

Budapest, 2024

Abstract

The deep learning revolution has fundamentally reshaped numerous fields, including robotics. However, as in other fields, certain challenges must be overcome to exploit the power of deep learning algorithms and create truly adaptive intelligent robots. The difficulty lies less in adult-level intelligence than in the skills of perception and mobility, also referred to as Moravec's paradox. In this context, the key issues are transferability and universality. This thesis addresses data-driven robotics, with a focus on transfer and curriculum learning. My main contributions are as follows.

Robots operating in unstructured environments need to effectively sense and interpret their surroundings. A major challenge for deep learning models in the field of robotics is the lack of domain-specific labelled data for various industrial applications. To bridge the reality gap, I developed a sim2real transfer learning method based on domain randomization for object detection (S2R-ObjDet), enabling automatic generation of labelled synthetic data. In addition, I propose the generalised confusion matrix (GCM) which addresses the limitations of the classical precision-recall-based metrics. I also introduce a public and annotated real-world dataset of industrial objects (InO-10-190) for evaluating sim2real object detection methods.

In object manipulation, it is essential to estimate not only object positions but also their poses. Thus, I propose two vision-based, multi-object grasp pose estimation models – the real-time MOGPE-RT and the high-precision MOGPE-HP – as well as the extension of the S2R-ObjDet method to pose estimation (S2R-PosEst). This framework provides an industrial tool for rapid data generation and model training while requiring minimal data from the target distribution.

Reinforcement learning – inspired by human learning – aims to offer a universal solution to various problems. Nevertheless, the field of robotics poses significant challenges. To facilitate the exploration of reinforcement learning robot agents, I propose a data exploitation curriculum learning method, called highlight experience replay (HiER). The experimental results demonstrate that HiER significantly improves the performance of the state-of-the-art, exhibiting stochastic dominance over them. To further enhance HiER, I introduce HiER+, which integrates an arbitrary data collection curriculum learning method for which I propose the easy2hard initial state entropy method (E2H-ISE).

Although the results presented in this thesis are my own, henceforth, I will use plural wording for stylistic purposes. The implementations, the qualitative results, the video presentations, and further materials are available on the project site: www.danielhorvath.eu/thesis.

Introduction

Deep learning (DL) is often regarded as the flagship of the modern artificial intelligence (AI) revolution. It has significantly transformed numerous fields including robotics. However, several challenges remain to be solved in order to fully harness the potential of DL algorithms and develop truly adaptive intelligent robots. This thesis tackles some of the key challenges in data-driven robotics, with a particular emphasis on transfer and curriculum learning. Due to space constraints, readers are encouraged to refer to the thesis for a more detailed introduction.

Sim2real knowledge transfer for object detection

Robots operating in unstructured environments must be capable of sensing and interpreting their surroundings. One of the main obstacles to deep-learning-based models in the field of robotics is the lack of domain-specific labelled data for different industrial applications. Thus, our first research question is the following: **How to transfer knowledge from simulation to the real world in the case of object detection?** Our theses regarding the first research question are as follows:

Thesis I: The synthetic images generated by our sim2real domain randomization method (S2R-ObjDet) enable object detection models to learn general representations of the objects, thereby bridging the gap between simulation and real-world environments.

We propose S2R-ObjDet, a domain-randomization-based sim2real synthetic data generation method for object detection. The 3D models of the given objects are loaded in the simulator, each with a random texture or monochromatic colour. Both the number and types of objects are randomised. Simulating gravitational force, the objects are dropped to a plane where they end up in one of their stable positions. The camera extrinsic and intrinsic parameters are set randomly with some constraints to ensure that the given objects are in the field of view. After an image is rendered, a post-processing method is applied to it involving multi-colour pepper-and-salt noise, gaussian blur, and optionally rectangular, circular, and line cutouts. The ground truth annotations of each object are automatically computed based on all points of the objects instead of the 8-points of the axis-aligned bounding boxes of the objects. This process is repeated until the required number of images for the training dataset is generated. S2R-ObjDet is capable of shrinking the reality gap between simulation and the real world to a satisfactory level, achieving 86.32% and 97.38% mAP_{50} scores respectively in the case of zero-shot and one-shot transfers, on our publicly available manually annotated InO-10-190 dataset, containing 190 real images of 920 object instances of 10 classes. The class selection was simultaneously based on different and similar objects in order to test the robustness of the model in terms of detecting different classes and differentiating between similar objects. Our solution fits industrial needs as the data generation process requires less than 0.5s per image enabling a fast training process. The training pipeline is presented in Fig. 1. This thesis is associated with [1].

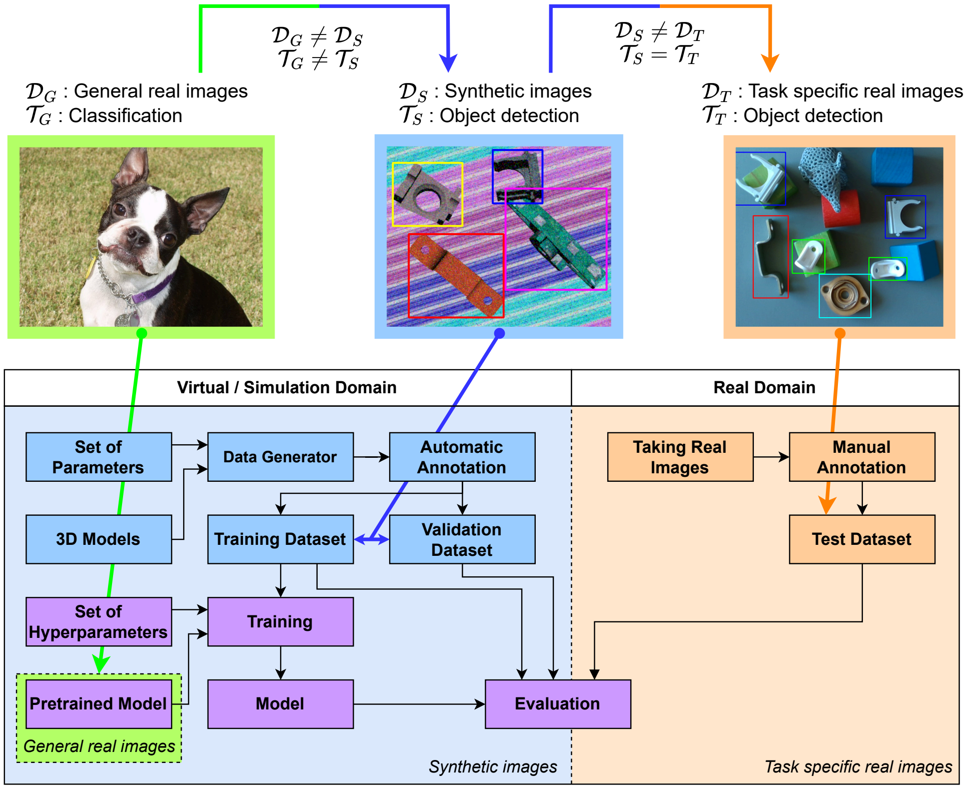
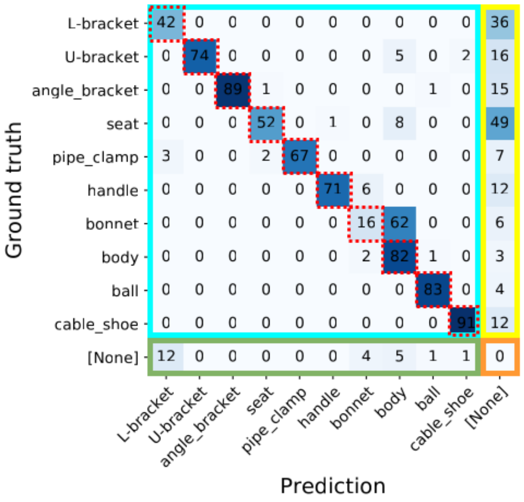

$\mathcal{D}_G \neq \mathcal{D}_S$
$\mathcal{T}_G \neq \mathcal{T}_S$
$\mathcal{D}_S \neq \mathcal{D}_T$
$\mathcal{T}_S = \mathcal{T}_T$
$\mathcal{D}_G$: General real images
$\mathcal{T}_G$: Classification
$\mathcal{D}_S$: Synthetic images
$\mathcal{T}_S$: Object detection
$\mathcal{D}_T$: Task specific real images
$\mathcal{T}_T$: Object detection
Virtual / Simulation Domain
Real Domain
Set of Parameters
Data Generator
Automatic Annotation
3D Models
Training Dataset
Validation Dataset
Set of Hyperparameters
Training
Pretrained Model
General real images
Model
Evaluation
Synthetic images
Taking Real Images
Manual Annotation
Test Dataset
Task specific real images

Figure 1: **Top.** Pipeline of knowledge transfer. **Bottom.** Flowchart diagram of our data generation, training, and evaluation process. The picture of the Boston bull is from ImageNet [9].

Thesis II: In object detection, misclassifications, false positives, and false negatives – factors not captured by traditional metrics – can be effectively quantified and evaluated using our generalised confusion matrix (GCM).

Our novel generalised confusion matrix (GCM) – depicted in Fig. 2 – is an adaptation of the classical confusion matrix to object detection. It addresses the limitations of the traditional precision-recall-based mAP and F_1 scores. Using the GCM, errors from misclassification, false positives, and false negatives can be effectively quantified and evaluated. Compared to the traditional confusion matrix $\boldsymbol{D} \in \mathbb{N}^{C \times C}$, where $C \in \mathbb{N}$ is the number of the classes, in our GCM $\boldsymbol{D}^{gen} \in \mathbb{N}^{C+1 \times C+1}$, one extra row and one extra column are added to the false positives and the false negatives cases. The correct detections are in the diagonal, $D^{gen}_{i,i}$, as in the case of the standard confusion matrix. $D^{gen}_{C+1,C+1} \doteq 0$. This thesis is associated with [1].

Our implementation, a quantitative and a qualitative evaluation, a video presentation, and further materials are available on the project site: www.danielhorvath.eu/sim2real.



	L-bracket	U-bracket	angle_bracket	seat	pipe_clamp	handle	bonnet	body	ball	cable_shoe	[None]
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

Figure 2: Generalised confusion matrix (GCM).

Sim2real grasp pose estimation

In the previous section, our sim2real domain randomization method was presented, focusing on object detection. Nevertheless, in object manipulation, it is essential to estimate not only object positions but also their orientations. Thus, our second research question is the following: **How to extend our S2R-ObjDet method to multi-object grasp pose estimation?** Our theses regarding the second research question are as follows:

Thesis III: Our novel two-stage multi-object grasp pose estimation methods – the real-time MOGPE-RT and the high-precision MOGPE-HP – enable a modular training approach for multi-object grasp pose estimation by utilizing sequential phases of object detection and class-specific orientation estimation.

We propose two vision-based, multi-object grasp pose estimation models – the real-time MOGPE-RT and the high-precision MOGPE-HP – depicted in Fig. 3. Both models are built upon two core components: an object detection model and an orientation estimation model. The output of the object detection model is $\boldsymbol{y} = \{(\boldsymbol{b}_i, c_i^{class}, p_i^{con}) \mid i = 1, 2, \ldots, N\}$*, where* $\boldsymbol{b}_i = [x_i, y_i, w_i, h_i] \in [0,1]^4$ *represents the axis-aligned bounding box of the* i^{th} *detection,* $c_i^{class} \in \mathbb{N}$ *is the class label of the* i^{th} *detection,* $p_i^{con} \in [0,1]$ *is the confidence score of the* i^{th} *detection, and* $N \in \mathbb{N}$ *is the number of detected objects. The detections with* $p_i^{con} < \tau_{con}$ *are filtered out, where* $\tau_{con} \in [0,1]$ *is the confidence threshold. The ROI cropping module extracts specific objects from the image and resizes them to the appropriate dimensions and shape. The class-specific orientation estimation models compute the* $sin(\theta_i)$ *and* $cos(\theta_i)$ *for all objects, where* $\theta_i \in [-\pi, \pi]$ *is the orientation angle. Then, with the atan2 function, the* θ_i *angles are computed which is the output of the MOGPE-RT model. In the case of the MOGPE-HP model, an additional local pattern-matching algorithm is incorporated, allowing for the estimation of a more precise* $\theta^* \in [-\pi, \pi]$ *at the expense of the extra computation. This thesis is associated with [4].*

Thesis IV: Our novel S2R-PosEst method facilitates rapid synthetic data generation for single-class orientation estimation models, effectively bridging the reality gap.

We propose S2R-PosEst, a sim2real domain randomization method for pose estimation, based on our S2R-ObjDet method. The 3D model of the given object is placed in the simulator and rotated around the z-axis – perpendicular to the plane where the object is placed – while random textures are added to the plane and to the object as well. All together, there are $n_{rot} = \lfloor \frac{2\pi}{\beta_{res}} \rfloor$ *rotations, where* $n_{rot} \in \mathbb{N}$ *is the number of rotation and* $\beta_{res} \in \mathbb{R}$ *is the resolution in radian. For each rotation, an image is taken and the label is automatically generated with it. The data generation requires 0.25–0.5s per image, making it suitable for industrial applications. This thesis is associated with [4].*

Our implementation, a quantitative and a qualitative evaluation, a video presentation, and further materials are available on the project site: www.danielhorvath.eu/mogpe.

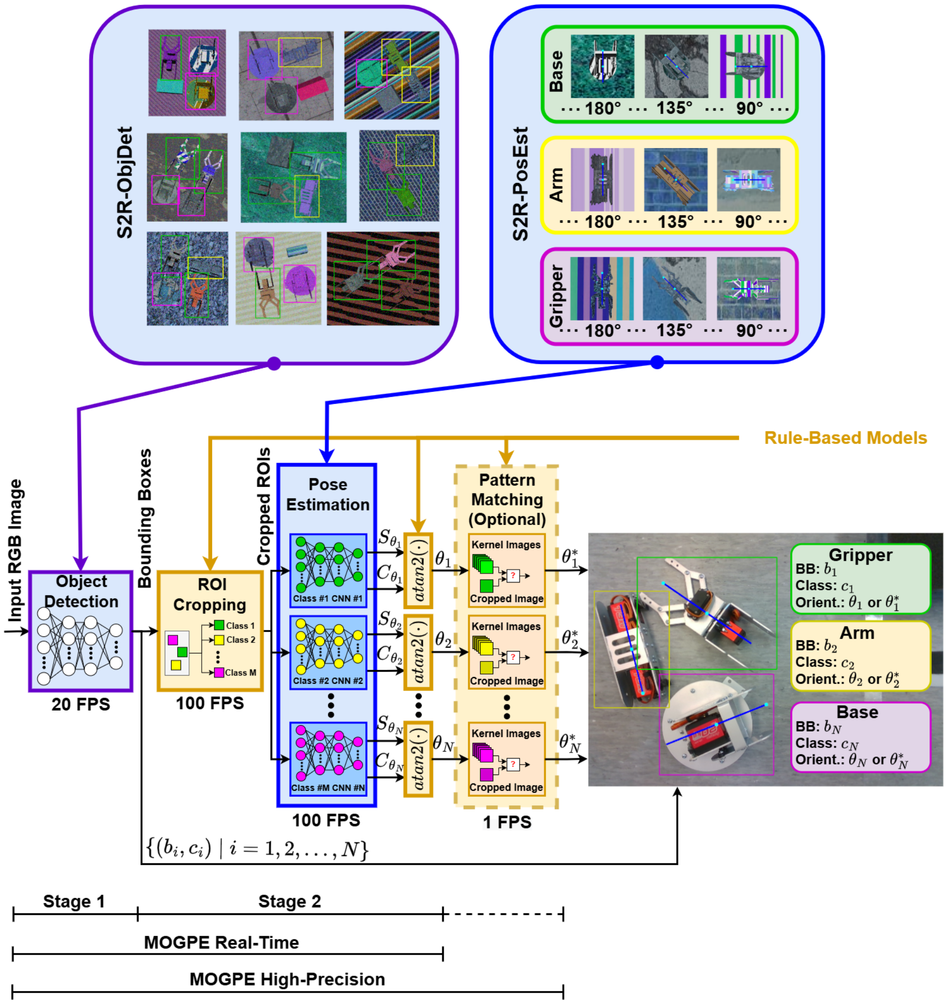


Figure 3: **Top.** Illustration of our S2N-ObjDet and S2N-PosEst methods. **Bottom.** Flowchart diagram of our multi-object grasp pose estimation (MOGPE) methods.

Highlight experience replay

In the previous sections, the main focus was on transferring knowledge from simulation to the real world in cases of supervised learning problems, namely object detection and pose estimation. Nonetheless, the endeavour for adaptive robots is coupled not only with transferability but universality as well. It is important to note that universal solutions are – by definition – easily transferable. An important building block in this attempt might be reinforcement learning (RL). Similarly to humans, RL algorithms learn from trial and error through interactions with the environment. Compared to supervised learning, RL is especially beneficial for robotic tasks that require a high level of dexterity. Nevertheless, the field of robotics poses significant challenges as the state and action spaces are continuous, and the reward function is predominantly sparse. Furthermore, on many occasions, the agent is devoid of access to any form of demonstration. Thus, our first research question is the following: **How to improve the training process of state-of-the-art reinforcement learning algorithms with curriculum learning?** Our theses regarding the third research question are as follows:

Thesis V: Our novel highlight experience replay (HiER) method enhances the training of reinforcement learning agents by separately storing and replaying the most relevant experiences, leading to a significant improvement in state-of-the-art performance.

Inspired by human learning, we propose HiER, the highlight experience replay method. A secondary experience replay buffer is created to store the most relevant transitions. At training, the transitions are sampled from both the standard experience replay buffer and the highlight experience replay buffer. It can be added to any off-policy RL agent and applied with or without the techniques of hindsight experience replay (HER) and prioritized experience replay (PER). HiER is depicted in Fig. 4 and detailed in Algorithm 1[a]. If only positive experiences are stored in its buffer, HiER can be viewed as a special, automatic demonstration generator as well. HiER is classified as a data exploitation or implicit curriculum learning method. HiER significantly improves the performance of RL baselines, having stochastic dominance over the state-of-the-art, validated on 8 tasks of three robotic benchmarks. This thesis is associated with [2].

---

[a] Due to the page limits, see Algorithm 1 in the Thesis.

Thesis VI: Our novel HiER+ approach enhances our highlight experience replay (HiER) method by increasing the availability of positive experiences – achieved through controlling task difficulty – particularly during the early stages of the training.

We propose HiER+ which is an enhancement of HiER with an arbitrary data collection (traditional) curriculum learning method. The overview of HiER+ is depicted in Fig.4 and detailed in Algorithm 2[a]*. Furthermore, as an example of the data collection CL method, we propose E2H-ISE, a universal, easy-to-implement easy2hard data collection CL method that requires minimal prior knowledge and controls the initial state-goal entropy (ISE) distribution* $\mathcal{H}(\mu_0)$ *which indirectly controls the task difficulty. Our experimental results show that HiER+ further improves HiER's performance. Moreover, HiER+ demonstrates stochastic dominance over HiER, based on the results from three robotic tasks of the Panda-Gym benchmark. This thesis is associated with [2].*

[a] Due to the page limits, see Algorithm 2 in the Thesis.

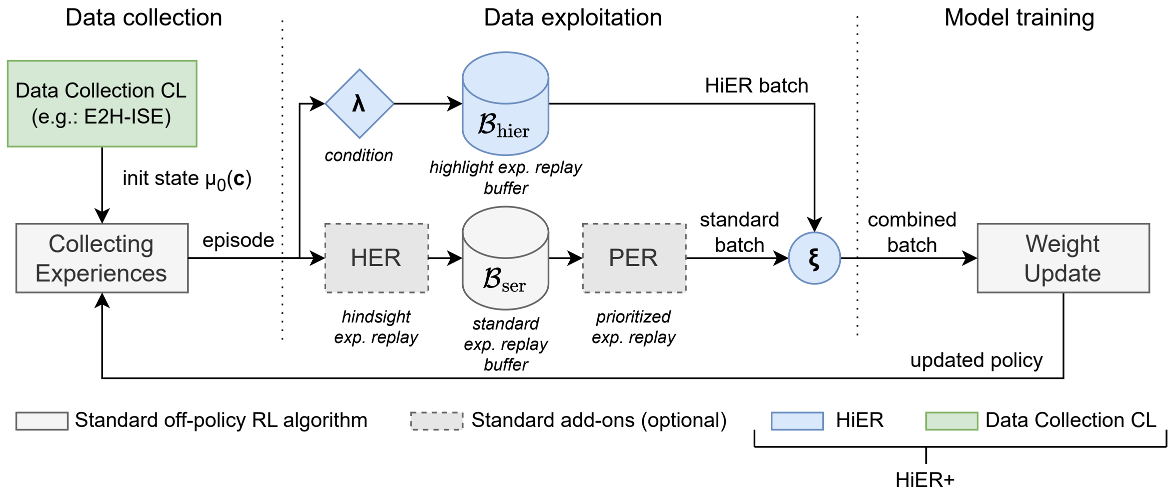

Data collection
Data exploitation
Model training
Data Collection CL (e.g.: E2H-ISE)
init state $\mu_0(\mathbf{c})$
Collecting Experiences
episode
λ
condition
$\mathcal{B}_{\text{hier}}$
highlight exp. replay buffer
HiER batch
HER
hindsight exp. replay
$\mathcal{B}_{\text{ser}}$
standard exp. replay buffer
PER
prioritized exp. replay
standard batch
ξ
combined batch
Weight Update
updated policy
Standard off-policy RL algorithm
Standard add-ons (optional)
HiER
Data Collection CL
HiER+

Figure 4: Overview of HiER and HiER+.

Our implementation, a quantitative and a qualitative evaluation, a video presentation, and further materials are available on the project site: www.danielhorvath.eu/hier

References

Author's journal papers

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

Robotics, vol. 39, no. 2, pp. 1225–1243, Apr. 2023, ISSN: 1941-0468. DOI: 10.1109/TRO.2022.3207619.

[2] **D. Horváth**, J. Bujalance Martín, F. Gábor Erdős, Z. Istenes, and F. Moutarde, "HiER: Highlight Experience Replay for Boosting Off-Policy Reinforcement Learning Agents," *IEEE Access*, vol. 12, pp. 100 102–100 119, Jul. 2024, ISSN: 2169-3536. DOI: 10.1109/ACCESS.2024.3427012.

[3] G. Erdős, K. Abai, R. Beregi, *et al.*, "Enabling Technologies for Autonomous Robotic Systems in Manufacturing," *Transactions of Nanjing University of Aeronautics and Astronautics*, vol. 41, no. 4, pp. 403–431, Aug. 2024, ISSN: 1005-1120. DOI: 10.16356/j.1005-1120.2024.04.001.

Author's conference papers

[4] **D. Horváth**, K. Bocsi, G. Erdős, and Z. Istenes, "Sim2Real Grasp Pose Estimation for Adaptive Robotic Applications," in *the 22nd IFAC World Congress*, ser. IFAC-PapersOnLine, vol. 56, 2023, pp. 5233–5239. DOI: 10.1016/j.ifacol.2023.10.121.

[5] G. Erdős, **D. Horváth**, and G. Horváth, "Visual Servo Guided Cyber-Physical Robotic Assembly Cell," in *the 17th IFAC Symposium on Information Control Problems in Manufacturing (INCOM)*, ser. IFAC-PapersOnLine, vol. 54, Jan. 2021, pp. 595–600. DOI: 10.1016/j.ifacol.2021.08.068.

[6] M. Hajós and **D. Horváth**, "Robotos Pakolási Feladat Megoldása Környezetérzékelés Segítségével," in *Nemzetközi Gépészeti Konferencia (OGÉT)*, Apr. 2020, pp. 305–308. [Online]. Available: https://ojs.emt.ro/oget/article/view/156.

[7] Z. Kemény, R. Beregi, J. Nacsa, C. Kardos, and **D. Horváth**, "Human–Robot Collaboration in the MTA SZTAKI Learning Factory Facility at Győr," in *the 8th CIRP Sponsored Conference on Learning Factories (CLF)*, ser. Procedia Manufacturing, vol. 23, Jan. 2018, pp. 105–110. DOI: 10.1016/j.promfg.2018.04.001.

[8] Z. Kemény, R. Beregi, J. Nacsa, C. Kardos, and **D. Horváth**, "Example of a Problem-to-Course Life Cycle in Layout and Process Planning at the MTA SZTAKI Learning Factories," in *the 9th Conference on Learning Factories (CLF)*, ser. Procedia Manufacturing, vol. 31, Jan. 2019, pp. 206–212. DOI: 10.1016/j.promfg.2019.03.033.

Most relevant references from the thesis

[9] J. Deng, W. Dong, R. Socher, *et al.*, "ImageNet: A Large-Scale Hierarchical Image Database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2009, pp. 248–255. DOI: 10.1109/CVPR.2009.5206848.

[10] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. The MIT Press, 2016, ISBN: 9780262035613.

[11] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction.* A Bradford Book, 2018, ISBN: 9780262039246.

[12] A. I. Károly, P. Galambos, J. Kuti, and I. J. Rudas, "Deep Learning in Robotics: Survey on Model Structures and Training Strategies," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 1, pp. 266–279, Jan. 2021, ISSN: 2168-2232. DOI: 10.1109/TSMC.2020.3018325.

[13] F. Zhuang, Z. Qi, K. Duan, *et al.*, "A Comprehensive Survey on Transfer Learning," *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, Jan. 2021, ISSN: 1558-2256. DOI: 10.1109/JPROC.2020.3004555.

[14] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor," *arXiv*, Aug. 2018. DOI: 10.48550/arXiv.1801.01290.

[15] X. Wang, Y. Chen, and W. Zhu, "A Survey on Curriculum Learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 9, pp. 4555–4576, Sep. 2022, ISSN: 1939-3539. DOI: 10.1109/TPAMI.2021.3069908.

[16] R. Portelas, C. Colas, L. Weng, K. Hofmann, and P.-Y. Oudeyer, "Automatic Curriculum Learning For Deep RL: A Short Survey," *arXiv*, May 2020. DOI: 10.48550/arXiv.2003.04664.

[17] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, "Curriculum Learning," in *Proceedings of the 26th Annual International Conference on Machine Learning (ICML)*, Association for Computing Machinery, Jun. 2009, pp. 41–48, ISBN: 9781605585161. DOI: 10.1145/1553374.1553380.

[18] J. Tobin, R. Fong, A. Ray, *et al.*, "Domain Randomization for Transferring Deep Neural Networks from Simulation to the Real World," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Sep. 2017, pp. 23–30. DOI: 10.1109/IROS.2017.8202133.

[19] A. Barisic, F. Petric, and S. Bogdan, "Sim2Air - Synthetic Aerial Dataset for UAV Monitoring," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 3757–3764, Apr. 2022, ISSN: 2377-3766. DOI: 10.1109/LRA.2022.3147337.

[20] T. Schaul, J. Quan, I. Antonoglou, and D. Silver, "Prioritized Experience Replay," *arXiv*, Feb. 2016. DOI: 10.48550/arXiv.1511.05952.

[21] M. Andrychowicz, F. Wolski, A. Ray, *et al.*, "Hindsight Experience Replay," in *Advances in Neural Information Processing Systems*, vol. 30, Curran Associates, Inc., Jul. 2017. DOI: 10.48550/arXiv.1707.01495.

[22] C. Florensa, D. Held, M. Wulfmeier, M. Zhang, and P. Abbeel, "Reverse Curriculum Generation for Reinforcement Learning," *arXiv*, Jul. 2018. DOI: 10.48550/arXiv.1707.05300.

[23] J. Tremblay, A. Prakash, D. Acuna, *et al.*, "Training Deep Networks with Synthetic Data: Bridging the Reality Gap by Domain Randomization," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Jun. 2018, pp. 1082–10 828. DOI: 10.1109/CVPRW.2018.00143.