Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Supplementary materials for

Amirfarhad FARHADI, Mitra MIRZAREZAEE, Arash SHARIFI, Mohammad TESHNEHLAB, 2024. Domain adaptation in reinforcement learning: a comprehensive and systematic study. *Front Inform Technol Electron Eng*, 25(11):1446-1465. https://doi.org/10.1631/FITEE.2300668

A reinforcement learning (RL) task that satisfies the Markov property is called a Markov decision process (MDP) (Sutton and Barto, 2018). An MDP is defined by several key equations:

State space S: a set of all possible environments or situations the agent can encounter.

Action space A: a set of all available choices the agent can make in each state.

Transition dynamics $T(s_{t+1} \mid s_t, a_t)$: a probability distribution depicting how the environment evolves to the next state s_{t+1} based on the current state s_t and chosen action a_t.

Reward function $R(s_t, a_t, s_{t+1})$: a numerical signal indicating the desirability of taking action a_t in state s_t and leading to the next state s_{t+1}.

Discount factor $\gamma \in [0, 1]$: a parameter balancing the importance of immediate rewards (γ=1) versus future rewards (γ closer to 0) (Arulkumaran et al., 2017).

In episodic tasks with a defined length T, reward accumulation over an entire episode can be described as

$$R=\sum_{t=0}^{T-1}\gamma^t r_{t+1}. \tag{S1}$$

The ultimate goal of RL is to find the optimal policy π that maximizes the expected return $\mathbb{E}[R|\pi]$ from all states:

$$\pi^*=\underset{\pi}{\operatorname{argmax}}\,\mathbb{E}[R|\pi]. \tag{S2}$$

For each interaction with the MDP, the agent begins with an initial state and then executes an action, which returns an outcome to guide the agent's actions (Anupong et al., 2023). An agent initiates the learning process by randomly performing an action that results in a specific environmental condition (Fig. S1). Next, the MDP, following the underlying transition dynamics, is transited to the next state. The agent collects rewards at a discounted rate because of its interaction with the MDP (Yin et al., 2022). The algorithm will learn a policy (i.e., an action–state relation) to choose the most optimal action in each situation and increase the cumulative reward (Bukhari et al., 2022).

Numerous RL algorithms have been introduced in recent years as the foundation for domain adaptation (DA) approaches. A summary of these algorithms is presented below.

1. Topological Q-learning. This algorithm guides the exploration to accomplish fast learning convergence based on the topological characteristics of the observable states of the environment in which the agent is operating (Hafez and Loo, 2015; Yin et al., 2023). It includes two main stages: task learning and exploration optimization. The instantaneous topological map (ITM) model creates a topological representation of the environment during the task learning stage to accelerate value function updates. During the exploration optimization

stage, an internal reward function is used to direct the exploration using the state values produced by the ITM nodes (Hafez and Kiong, 2014; Li HQ et al., 2023). Therefore, this algorithm is intended to provide directed exploration through an intrinsic or internal reward method. Since the guided exploration has a clear purpose, it appears preferable to any random exploration approach. The state–action value function of the Q-learning agent, represented as $Q(s, a)$, is updated in accordance with the Bellman equation. Let s represent the current state, a the action taken, r the received reward, and $\acute{s}$ the next state. The value of a state, $V(\acute{s})$, is determined by identifying the action that yields the highest expected long-term return from that state. This is achieved by considering the state–action values. The value of the next state $\acute{s}$, $V(\acute{s})$, is computed by selecting the maximum over all possible actions of the state–action values.

$$Q(s, a)=Q(s, a)+\alpha(r + V(\acute{s}) - Q(s, a)). \tag{S3}$$

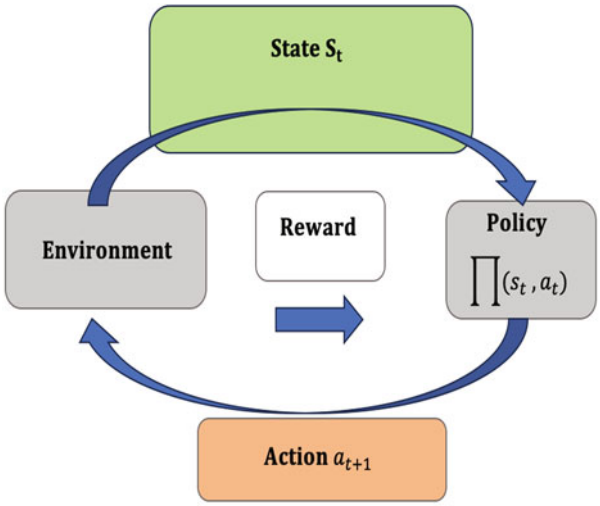

Fig. S1 Interaction between the environment and agents in reinforcement learning

In the iterative interaction between the RL agent and its environment, each newly sensed state expands the topological map with a dedicated node. Upon taking action a at state s_{n-1} (Fig. S2), the value estimate $V(s_{n-1})$ is updated as per Eq. (S3). This update then propagates through the immediate neighborhood $N(s_{n-1})$ via existing edges, modifying the value estimates of connected nodes according to the same equation.

2. Epoch-incremental. This type of learning combines epoch and incremental learning methods that enable policy modification in both modes. The epoch mode is similar to a breadth-first search (BFS) algorithm in numerous aspects. There are two implementations of epoch-incremental algorithms in the literature (Luo et al., 2022). The first involves using the agent experience to determine the parameters of a learning system. In this manner, the distances from the examined nodes to the terminal state are estimated, and the policy with the least distance between any initial state and the terminal state is considered the optimum policy. Supervised learning in the epoch mode is used to restrict the number of costly experiments using real experiments during the initialization of the value function. The second implementation, which combines the algorithm in the epoch and incremental mode, relies on directing the RL in the incremental mode (Zajdel, 2018).

3. Multi-scale. This uses mathematical functions to create an abstract of the state-space graph. Action selection on the reduced abstraction map is carried out using multi-scale Softmax selection. This could be considered a simplified mathematical modeling of real-world scenarios (Ma et al., 2023). In this regard, as with any simplified model, there is a concern about oversimplification and omitting details, which may result in lower performance in the real world. However, the benefits of simplified models cannot be ignored. The key trick in

these models is balancing oversimplification and keeping abstraction (Fu et al., 2023).

4. Hierarchical. This approach is an innovative strategy for expanding the capabilities of traditional RL algorithms to handle more complex tasks. It decomposes a larger goal into a hierarchy of sub-goals (Gao et al., 2023). Each sub-goal can be further subdivided into a set of subtasks, with primitive operations occurring at the lowest level of tasks. The hierarchical approach is similar to the work breakdown structure (WBS) used in various software project management methods (Nachum et al., 2018).

5. Deep RL. Deep RL combines RL and deep learning, providing high-level data abstraction via multi-layer graph processing. It enables the determination of data relevant to predefined objectives. The algorithm is based on two fundamental concepts: (1) an experience replay mechanism that eliminates correlations among consecutive observations and (2) an iterative update mechanism that adjusts the Q-values to make them closer to the goal value (François-Lavet et al., 2018).

6. Temporal-difference learning. This method, which combines notions from Monte Carlo estimation and dynamic programming, plays an important role in RL. A basic aspect of this method is that it gradually acquires testable and predictive knowledge about the environment. The learned values solve queries regarding how a signal accumulates over time in response to a certain behavior. In control tasks, this signal indicates how many rewards an agent is likely to gain if it acts greedily relative to its current predictions (De Asis et al., 2020).

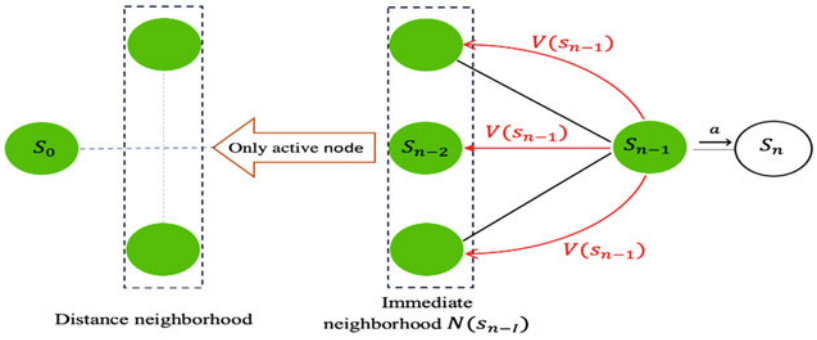

Fig. S2 Backpropagation in topological Q-learning

Transfer learning (TL) is a promising machine learning technique for dealing with inadequate training data that has recently gained much research interest (Sadr et al., 2021; Farhadi and Sharifi, 2024). The term "learning scenario" refers to training a model on a source task or domain and assessing it on a target task or domain, where the domains or tasks may vary (Li S et al., 2018). For example, a synthetic traffic sign dataset that is simple to create may be used in training a model that will be used to identify actual traffic signs. Furthermore, a hand-written digit dataset may be used to train a model to recognize house numbers. In these instances, the training and evaluation datasets are related but distinct. TL can be used to improve the accuracy of target data if the source and target data differ (Li DM, 2022).

Two keywords, domain and task, were developed by Pan and Yang (2010) to aid in classifying different TL techniques. Feature spaces and marginal probability distributions form domains. A task consists of labels and models derived from training. Therefore, TL problems may be classified as transferring information from one domain to another or transferring knowledge between tasks. Accordingly, feature space changes or marginal distribution changes may cause domain changes. When categorizing records through text mining, language changes may cause a change in the feature space, whereas any change in the document subjects may cause changes in marginal probability distributions (Ghahfarrokhi et al., 2023). Moreover, task changes may result from changes in the label space or objective predictive functions. Regarding document categorization, the label space may change by increasing or decreasing the number of classes (Liu X et al., 2023).

Pan and Yang (2010) proposed three categories for classifying TL algorithms (inductive, transductive, and unsupervised) depending on whether the domain or task differs from the source and the target. In inductive TL,

some labeled target data are needed, sources and targets are distinct, and domains may or may not vary. Transductive TL requires labeled sources and unlabeled targets. The tasks stay the same in this case, but the domains are different. In unsupervised TL, the tasks vary from those in inductive TL, and neither the target nor the source domains need to be labeled (Jafari et al., 2023). Over the last decade, several strategies have been developed to overcome TL issues, such as multi-view, multi-domain, and multi-task learning, as described below.

1. Multi-view learning. It concerns machine learning issues and includes several data views. A view represents a unique collection of available features (Liu QY et al., 2023). Multiple views are shown intuitively because a video object may be represented from two distinct perspectives: (1) the picture signal and (2) the audio signal. Multi-view learning defines an item from various perspectives, resulting in abundant information (Li DM et al., 2022). Learners' performance can be enhanced by carefully considering information from all views. Multi-view learning involves various methods, including subspace, multi-kernel, and co-training (Zhao J et al., 2017). Recently, various TL approaches have been proposed based on multi-view techniques. For instance, Feuz and Cook (2017) developed a multi-view TL method for activity learning that facilitates the transfer of activity knowledge across diverse sensor platforms. Yang P and Gao (2013) used multi-view information to facilitate knowledge transfer across multiple domains, and Zhang D et al. (2011) developed a multi-view TL model that ensures consistency among various views.

2. Multi-domain learning. It aims to disseminate knowledge about the same topic across multiple contextual domains. This method addresses the problem of learning several models generated by a common deep architecture customized to fulfill a task in a particular domain. Efforts have recently been made to adapt deep architectures to novel domains and tasks. Earlier studies had addressed simple solutions, including fine-tuning existing pretrained models, requiring multiple specialized models, and incurring catastrophic forgetting. Recent research has investigated the issue and shown how to extend the abilities of current deep architectures by incorporating parameters tailored to specific tasks (Sun et al., 2018). Rebuffi et al. (2017) proposed residual adapters, including a design for residual blocks that incorporate task-specific components. Rebuffi et al. (2018) presented an architecture in which the topology of the adapters is parallel rather than series. Mallya and Lazebnik (2018) examined weight-based pruning to combine multiple tasks into a single neural network.

3. Multi-task learning. It uses domain-specific information from related tasks to train different classifiers that benefit from one another cooperatively. Specifically, multi-task learning reinforces each task, taking advantage of the interconnectedness between them (inter-task relevancy and cross-task relationship) and improving their generalizability (Huang et al., 2023). In multi-task learning, knowledge is transferred across related domains, whereas in TL, knowledge is transferred across multiple relevant tasks. Multi-task learning prioritizes the tasks equally, while TL prioritizes the target tasks over the sources (Zhao X et al., 2022). Nonetheless, overlaps and connections exist between multi-task learning and TL. Both aim to enhance learners' performance by transferring knowledge, and they use common modeling methodologies like parameter sharing and feature transformation (Kouw and Loog, 2019). Several studies have adopted both multi-task learning and TL methods in their innovation. For instance, Zhang W et al. (2016) used multi-task and TL approaches to analyze biological images, and Liu AA et al. (2019) introduced a scheme for recognizing human actions based on multi-source TL and multi-task learning.

Table S1 Details of domain adaptation methods in the context of reinforcement learning

Application domain	Reference	Function	Algorithm	Experimental platform	Dataset	Pros	Cons
RL and dialogue systems	Su et al. (2017)	Prediction	Policy gradient	Python multi-domain statistical dialogue system toolkit (PyDial)	N/A	The effective use of demonstration data to improve policy learning at an early stage.	The neural network architectures used in these methods may not handle uncertainty effectively, leading to lower performance in noisy environments than other RL techniques.
	Yang M et al. (2018)	Prediction	Policy gradient	N/A	The original SQuAD dataset comprises 87,636 training instances and 10,600 development instances	It outperforms both supervised and unsupervised learning techniques, including dual learning and adversarial DA.	The interpretation of emojis can be subjective and context-dependent, which can lead to misunderstandings or miscommunications if the sender and receiver have different interpretations or expectations.
	Patel et al. (2018)	Prediction	SVM and DQN	Python	Office-31	The learned policies outperform baselines, but they fail to close the gap towards recent unsupervised adaptation techniques.	It uses fixed representations for both the source and target domain samples, which may limit its performance.
	Shoeleh and Asadpour (2020)	Mapping	Graph	N/A	N/A	It is capable of establishing relations between the tasks, successfully transferring the acquired abstract skills, and enhancing the agent's performance in the target task by applying the gained skills.	The effectiveness of the method depends heavily on the similarity between the source and target tasks. If the tasks are too dissimilar, the method may not be effective or may even lead to negative transfer.
Data valuation in machine learning	Yoon et al. (2020)	Prediction	Regression and Bayes	N/A	Twelve public datasets, including two language datasets (SMS spam and Email spam), seven image datasets (CIFAR-10, CIFAR-100, Fashion-MNIST, Flower, USPS, MNIST, and HAM 10000), and three tabular datasets (Rossmann Store Sales, Adult, and Blog)	It can offer high-quality training data ranking in a computationally efficient manner, which is beneficial for DA, corrupted sample discovery, and robust learning.	The accuracy of the DVE model used in DVRL is dependent on the quality and representativeness of the training data, which can be a challenge in some cases.
	Liu BY et al. (2020)	Prediction	DQN	Real lab environment	DomainNet, Office-3, and Office-Hom	The incorporation of target margin loss during base model training can enhance the discriminability of the model, which can lead to better performance on benchmark datasets.	The effectiveness of the approach may depend on the quality and representativeness of the training data, which can be a challenge in some cases.
Sim2real transfer	Chen HX et al. (2021)	Mapping	PPO	MuJoCo	It includes 600 trajectories collected by a sub-optimal policy	The use of a learned mapping function from images in the target domain and states in the source domain can enable policies trained on states in the source domain to be applied directly in the target domain of images, which can reduce the need for expensive and time-consuming simulators.	The approach may not be suitable for all types of sim2real transfer problems, as it specifically focuses on learning a mapping function from images in the target domain and states in the source domain.

Table S1 (continued)

Application domain	Reference	Function	Algorithm	Experimental platform	Dataset	Pros	Cons
	Truong et al. (2021)	Prediction	DD-PPO	Real lab environment	*N* unpaired images	The bi-directional DA strategy leverages both simulation and reality differences to improve the generalization of RL policies and expedite learning. The real2sim strategy separates the sensor adaptation module from the policy training, which can reduce an extra bottleneck during the RL policy training process.	The approach may not be suitable for all types of sim2real transfer problems, as it specifically focuses on the task of point goal navigation using egocentric RGB-D observations.
Visual control and robotics	Zhang H et al. (2022)	Mapping	A3C	Gazebo	N/A	The approach addresses the reality gap by converting real-world image streams back to the synthetic domain during the deployment stage, which can make the robot feel at home and improve the transfer of deep RL policies acquired in simulated environments to the real-world domains for visual control tasks.	The approach may not be suitable for all types of visual control tasks, as it specifically focuses on converting real-world image streams back to the synthetic domain for deep RL policies acquired in simulated environments.
	Li SD et al. (2020)	Mapping	PPO	Habitat	Gibson	Experimental results demonstrate that the approach outperforms baselines in transferring navigation policies for various tasks between two simulation domains and from simulation to the real world, which suggests its potential for real-world applications.	The approach may require additional computational resources for the image transformation process, which can increase the training and deployment time.
Text and language processing	Yang M et al. (2018)	Sequence-to-sequence	Policy gradient	N/A	Twitter and Sina	The model learns the human responding style from vast amounts of generic data and fine-tunes the model using a small amount of customized data to create customized conversations, which can improve the quality of the responses for various users.	The model may require additional computational resources for the training and optimization process, which can increase the training and deployment time.
	Jeong et al. (2020)	Prediction	DDPG	MuJoCo	N/A	Compared to domain randomization, only 5 h of real robot data are used for adaptation.	The approach is specifically designed for cube stacking from visual observations, which limits its applicability to other robotic manipulation tasks. The approach involves training a dynamic model and optimizing the RL policy network, which can be computationally expensive and time-consuming.
	Liu MF et al. (2019)	Prediction	Policy gradient	N/A	SANCL	This approach is efficient at data selection and representation and generalizable to accommodate a variety of NLP tasks.	The experimental findings presented in the paper are limited to sentiment analysis, dependency parsing, and part-of-speech tagging. It is unclear how well the proposed approach would perform on other NLP tasks or in different domains.

Table S1 (continued)

Application domain	Reference	Function	Algorithm	Experimental platform	Dataset	Pros	Cons
	Chen J et al. (2022)	Auto encoding	Deep Q-learning	N/A	Office-31, Office+Caltech-10, and Caltech-Office	By evaluating the importance of the source items to the target domain, the agent is guided to learn appropriate selection policies.	The deep Q-learning algorithm used in the approach requires significant computational resources, which may limit its scalability to larger datasets or more complex tasks.
Representation learning and disentangled representations	Higgins et al. (2017)	Mapping	DQN, A3C, and episodic control	MuJoCo	N/A	It can obtain source policies that are robust to various domain changes without having access to the target domain.	The performance of the DARLA approach can be sensitive to the choice of hyperparameters, which may require careful tuning.
	Carr et al. (2019)	Auto encoding	A2C	ALE	N/A	They have shown how a basic approach can assist learning in RL, even when the last layer needs to be relearned due to changes in the action space and task.	It may struggle to adapt to highly complex or non-stationary environments, where the distribution of data may change rapidly or unpredictably.
Semantic representation learning	Dong et al. (2020)	Prediction	Regression	FCN-8s	NTHU, SYNTHIA, GTA, and Cityscapes	It achieves better performance in mitigating significant distribution shifts for classes with different appearances across multiple datasets by investigating transferable representations.	The authors evaluated their approach on only a few datasets, which may limit the generalizability of their findings.

References

Anupong W, Mehbodniya A, Webber JL, et al., 2023. Deep learning algorithms were used to generate photovoltaic renewable energy in saline water analysis via an oxidation process. *Water Reuse*, 13(1):68-81. https://doi.org/10.2166/wrd.2023.071

Arulkumaran K, Deisenroth MP, Brundage M, et al., 2017. A brief survey of deep reinforcement learning. https://doi.org/10.48550/arXiv.1708.05866

Bukhari SNH, Webber J, Mehbodniya A, 2022. Decision tree based ensemble machine learning model for the prediction of Zika virus T-cell epitopes as potential vaccine candidates. *Sci Rep*, 12:7810. https://doi.org/10.1038/s41598-022-11731-6

Carr T, Chli M, Vogiatzis G, 2019. Domain adaptation for reinforcement learning on the Atari. 18th Int Conf on Autonomous Agents and MultiAgent Systems, p.1859-1861.

Chen J, Wu XX, Duan LX, et al., 2022. Domain adversarial reinforcement learning for partial domain adaptation. *IEEE Trans Neur Netw Learn Syst*, 33(2):539-553. https://doi.org/10.1109/TNNLS.2020.3028078

Chen XH, Jiang S, Xu F, et al., 2021. Cross-modal domain adaptation for cost-efficient visual reinforcement learning. 35th Conf on Neural Information Processing Systems, p.12520-12532.

De Asis K, Chan A, Pitis S, et al., 2020. Fixed-horizon temporal difference methods for stable reinforcement learning. Proc AAAI Conf on Artificial Intelligence.

Dong JH, Cong Y, Sun G, et al., 2020. CSCL: critical semantic-consistent learning for unsupervised domain adaptation. 16th European Conf on Computer Vision, p.745-762. https://doi.org/10.1007/978-3-030-58598-3_44

Farhadi A, Sharifi A, 2024. Leveraging meta-learning to improve unsupervised domain adaptation. *Comput J*, 67(5):1838-1850. https://doi.org/10.1093/comjnl/bxad104

Feuz KD, Cook DJ, 2017. Collegial activity learning between heterogeneous sensors. *Knowl Inform Syst*, 53:337-364.

François-Lavet V, Henderson P, Islam R, et al., 2018. An introduction to deep reinforcement learning. https://doi.org/10.48550/arXiv.1811.12560

Fu C, Yuan H, Xu H, et al., 2023. TMSO-Net: texture adaptive multi-scale observation for light field image depth estimation. *J Vis Commun Image Represent*, 90:103731. https://doi.org/10.1016/j.jvcir.2022.103731

Gao J, Wu DZ, Yin F, et al., 2023. MetaLoc: learning to learn wireless localization. *IEEE J Sel Areas Commun*, 41(12):3831-3847. https://doi.org/10.1109/JSAC.2023.3322766

Ghahfarrokhi SS, Khodadadi H, Ghadiri H, et al., 2023. Malignant melanoma diagnosis applying a machine learning method based on the combination of nonlinear and texture features. *Biomed Signal Process Contr*, 80:104300. https://doi.org/10.1016/j.bspc.2022.104300

Hafez MB, Kiong LC, 2014. Curiosity-based topological reinforcement learning. IEEE Int Conf on Systems, Man, and Cybernetics.

Hafez MB, Loo CK, 2015. Topological Q-learning with internally guided exploration for mobile robot navigation. *Neur Comput Appl*, 26:1939-1954. https://doi.org/10.1007/s00521-015-1861-8

Higgins I, Pal A, Rusu A, et al., 2017. DARLA: improving zero-shot transfer in reinforcement learning. 34th Int Conf on Machine Learning, p.1480-1490.

Huang W, Wang X, Zhang J, et al., 2023. Improvement of blueberry freshness prediction based on machine learning and multi-source sensing in the cold chain logistics. *Food Contr*, 145:109496. https://doi.org/10.1016/j.foodcont.2022.109496

Jafari BM, Luo X, Jafari A, 2023.. Unsupervised keyword extraction for hashtag recommendation in social media. Int FLAIRS Conf Proc.

Jeong R, Aytar Y, Khosid D, et al., 2020. Self-supervised sim-to-real adaptation for visual robotic manipulation. IEEE Int Conf on Robotics and Automation, p.2718-2724. https://doi.org/10.1109/10.1109/ICRA40945.2020.9197326

Kouw WM, Loog M, 2019. A review of domain adaptation without target labels. *IEEE Trans Patt Anal Mach Intell*, 43(3):766-785. https://doi.org/10.1109/TPAMI.2019.2945942

Li DM, 2022. Machine learning based preschool education quality assessment system. *Mob Inform Syst*, 2022:2862518. https://doi.org/10.1155/2022/2862518

Li DM, Dai X, Wang J, et al., 2022. Evaluation of college students' classroom learning effect based on the neural network algorithm. *Mob Inform Syst*, 2022:7772620. https://doi.org/10.1155/2022/7772620

Li HQ, Huang J, Cao Z, et al., 2023. Stochastic pedestrain avoidance for autonomous vehicle using hybrid reinforcement learning. *Front Inform Technol Electron Eng*, 24(1):131-140. https://doi.org/10.1631/FITEE.2200128

Li S, Song SJ, Cheng W, 2018. Layer-wise domain correction for unsupervised domain adaptation. *Front Inform Technol Electron Eng*, 19(1):91-103. https://doi.org/10.1631/FITEE.1700774

Li SD, Chaplot DS, Tsai YHH, et al., 2020. Unsupervised domain adaptation for visual navigation. https://doi.org/10.48550/arXiv.2010.14543

Liu AA, Xu N, Nie WZ, et al., 2019. Multi-domain and multi-task learning for human action recognition. *IEEE Trans Image Process*, 28(2):853-867. https://doi.org/10.1109/TIP.2018.2872879

Liu BY, Guo YH, Ye JP, et al., 2020. Selective pseudo-labeling with reinforcement learning for semi-supervised domain adaptation. 32nd British Machine Vision Conf, p.299.

Liu MF, Song Y, Zou HB, et al., 2019. Reinforced training data selection for domain adaptation. Proc 57th Annual Meeting of the

Association for Computational Linguistics, p.1957-1968. https://doi.org/10.18653/v1/P19-1189
Liu QY, Li DQ, Tang XS, et al., 2023. Predictive models for seismic source parameters based on machine learning and general orthogonal regression approaches. *Bull Seismol Soc Am*, 113:2363-2376. https://doi.org/10.1785/0120230069
Liu X, Shi T, Zhou G, et al., 2023. Emotion classification for short texts: an improved multi-label method. *Human Soc Sci Commun*, 10:306. https://doi.org/10.1057/s41599-023-01816-6
Luo J, Wang G, Li G, et al., 2022. Transport infrastructure connectivity and conflict resolution: a machine learning analysis. *Neur Comput Appl*, 34(9):6585-6601.
Ma B, Liu Z, Dang Q, et al., 2023. Deep reinforcement learning of UAV tracking control under wind disturbances environments. *IEEE Trans Instrum Meas*, 72:2510913. https://doi.org/10.1109/TIM.2023.3265741
Mallya A, Lazebnik S, 2018. Packnet: adding multiple tasks to a single network by iterative pruning. Proc IEEE Conf on Computer Vision and Pattern Recognition.
Nachum O, Gu SX, Lee H, et al., 2018. Data-efficient hierarchical reinforcement learning. 32nd Int Conf on Neural Information Processing Systems, p.3307-3317.
Pan SJ, Yang Q, 2010. A survey on transfer learning. *IEEE Trans Knowl Data Eng*, 22(10):1345-1359. https://doi.org/10.1109/TKDE.2009.191
Patel Y, Chitta K, Jasani B, 2018. Learning sampling policies for domain adaptation. https://doi.org/10.48550/arXiv.1805.07641
Rebuffi SA, Bilen H, Vedaldi A, 2017. Learning multiple visual domains with residual adapters. https://doi.org/10.48550/arXiv.1705.08045
Rebuffi SA, Bilen H, Vedaldi A, 2018. Efficient parametrization of multi-domain deep neural networks. Proc IEEE Conf on Computer Vision and Pattern Recognition.
Sadr H, Pedram MM, Teshnehlab M, 2021. Convolutional neural network equipped with attention mechanism and transfer learning for enhancing performance of sentiment analysis. *J AI Data Mining*, 9(2):141-151. https://doi.org/10.22044/jadm.2021.9618.2100
Shoeleh F, Asadpour M, 2020. Skill based transfer learning with domain adaptation for continuous reinforcement learning domains. *Appl Intell*, 50(2):502-518. https://doi.org/10.1007/s10489-019-01527-z
Su PH, Budzianowski P, Ultes S, et al., 2017. Sample-efficient actor-critic reinforcement learning with supervised data for dialogue management. 18th Annual SIGdial Meeting on Discourse and Dialogue, p.147-157. https://doi.org/10.18653/v1/W17-5518
Sun G, Li Y, Liao D, et al., 2018. Service function chain orchestration across multiple domains: a full mesh aggregation approach. *IEEE Trans Netw Serv Manag*, 15:1175-1191. https://doi.org/10.1109/TNSM.2018.2861717
Sutton RS, Barto AG, 2018. Reinforcement Learning: an Introduction (2nd Ed.). The MIT Press.
Truong J, Chernova S, Batra D, 2021. Bi-directional domain adaptation for Sim2Real transfer of embodied navigation agents. *IEEE Robot Autom Lett*, 6(2):2634-2641. https://doi.org/10.1109/LRA.2021.3062303
Yang M, Tu W, Qu Q, et al., 2018. Personalized response generation by dual-learning based domain adaptation. *Neur Netw*, 103:72-82. https://doi.org/10.1016/j.neunet.2018.03.009
Yang P, Gao W, 2013. Multi-view discriminant transfer learning. 23rd Int Joint Conf on Artificial Intelligence.
Yin Y, Guo Y, Su Q, et al., 2022. Task allocation of multiple unmanned aerial vehicles based on deep transfer reinforcement learning. *Drones*, 6:215. https://doi.org/10.3390/drones6080215
Yin Y, Zhang R, Su Q, 2023. Threat assessment of aerial targets based on improved GRA-TOPSIS method and three-way decisions. *Math Biosci Eng*, 20(7):13250-13266.
Yoon J, Arik S, Pfister T, 2020. Data valuation using reinforcement learning. 37th Int Conf on Machine Learning, p.10842-10851.
Zajdel R, 2018. Epoch-incremental Dyna-learning and prioritized sweeping algorithms. *Neurocomputing*, 319:13-20. https://doi.org/10.1016/j.neucom.2018.08.068
Zhang D, He J, Liu Y, et al., 2011. Multi-view transfer learning with a large margin approach. Proc 17th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining.
Zhang H, Luo G, Li J, et al., 2022. C2FDA: coarse-to-fine domain adaptation for traffic object detection. *IEEE Trans Intell Transp Syst*, 23:12633-12647. https://doi.org/10.1109/TITS.2021.3115823
Zhang W, Li R, Zeng T, et al., 2016. Deep model based transfer and multi-task learning for biological image analysis. *IEEE Trans Big Data*, 6:322-333. https://doi.org/10.1109/TBDATA.2016.2573280
Zhao J, Xie X, Xu X, et al., 2017. Multi-view learning overview: recent progress and new challenges. *Inform Fusion*, 38:43-54.
Zhao X, Yang M, Qu Q, et al., 2022. Exploring privileged features for relation extraction with contrastive student-teacher learning. *IEEE Trans Knowl Data Eng*, 35(8):7953-7965. https://doi.org/10.1109/TKDE.2022.3161584