CIRP Annals - Manufacturing Technology 73 (2024) 341–344

Contents lists available at ScienceDirect

CIRP Annals - Manufacturing Technology

journal homepage: https://www.editorialmanager.com/CIRP/default.aspx

A vision-language-guided and deep reinforcement learning-enabled approach for unstructured human-robot collaborative manufacturing task fulfilment

Pai Zheng[a], Chengxi Li[a], Junming Fan[a], Lihui Wang (1)[b,*]

[a] Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hong Kong Special Administrative Region (HKSAR), China
[b] Department of Production Engineering, KTH Royal Institute of Technology, Sweden

ARTICLE INFO

Article history:
Available online 17 April 2024

Keywords:
Human-robot collaboration
Manufacturing system
Human-guided robot learning

ABSTRACT

Human-Robot Collaboration (HRC) has emerged as a pivot in contemporary human-centric smart manufacturing scenarios. However, the fulfilment of HRC tasks in unstructured scenes brings many challenges to be overcome. In this work, mixed reality head-mounted display is modelled as an effective data collection, communication, and state representation interface/tool for HRC task settings. By integrating vision-language cues with large language model, a vision-language-guided HRC task planning approach is firstly proposed. Then, a deep reinforcement learning-enabled mobile manipulator motion control policy is generated to fulfil HRC task primitives. Its feasibility is demonstrated in several HRC unstructured manufacturing tasks with comparative results.

1. Introduction

In line with the evolving human-centricity trend for Industry 5.0, Human-Robot Collaboration (HRC) has emerged as a prevailing manufacturing paradigm. It simultaneously takes the advantages of human cognitive flexibility and robotic automation capability to achieve better production efficiency [1]. Nevertheless, existing robotic cells have encountered a significant hurdle in their ability to handle unstructured manufacturing tasks, which require a high level of cognition in mass personalization. These tasks typically encompass intricate product assembly, disassembly, and inspection processes, characterized by the absence of predefined structures or instructions necessitating swift adaptation. As a result, extant HRC systems face difficulties in performing effectively in such unstructured scenarios, leading to inefficiencies and limitations in executing tailored manufacturing tasks.

On one hand, to strengthen robotic cognition capabilities, multimodal intelligence based HRC approaches have been widely investigated in the manufacturing domain. Among them, vision-based ones often serve as the main perception channel, owing to its cost-effectiveness [2]. Meanwhile, the very recent breakthrough of large language models (LLMs) has also attracted great interest in the natural language guided HRC applications. Nevertheless, as an emerging topic, how to align the visual and linguistic cues, and comprehend the underlying semantics via LLMs to jointly facilitate HRC in the manufacturing scenarios is yet to be much explored.

On the other hand, robot skills (e.g., tasks, working steps, trajectories) are normally pre-programmed in a fixed workstation for automation efficiency. However, when it comes to an unstructured manufacturing scene and/or unpredictable human interventions, existing solutions fail to adapt effectively. To meet such demand, next-generation robot manipulators should be well-equipped with human-guided learning capabilities to collaborate and even co-evolve with human operators effectively.

To address these two issues, this work introduces a vision-language-guided deep reinforcement learning (DRL)-enabled planning approach for unstructured HRC manufacturing task fulfilment via Mixed-Reality Head-Mounted Display (MR-HMD). Firstly, the potential of MR-HMD is thoroughly explored, transforming it into an efficient tool for data collection, communication, and state representation in HRC. This involves capturing structured scene information, language guidance instructions, and raw image streams. Secondly, a vision-language understanding model is developed based on acquired language and image data, enabling adaptive HRC task planning. Specifically, a vision-language-guided target object segmentation model is devised to localize robotic action goals, while an LLM-based robotic task planning module generates action plans using language commands. Finally, a time-aware DRL-based whole-body motion planning policy is proposed, utilizing planned tasks, acquired vectors, and image sequences to successfully complete diverse HRC tasks, while ensuring human safety. The rest of this paper is organized as follows: Section 2 gives an overall systematic methodology encompassing the vision-language guidance and DRL-based whole-body motion planning via MR-HMD. Section 3 provides a demonstrative case study to validate its feasibility with comparative analysis. At last, Section 4 summarizes the major findings and highlights the potential future works.

2. Methodology

To fulfil various HRC manufacturing tasks in unstructured scenes, the proposed system framework is depicted in Fig. 1. When completing various HRC tasks, human operators wearing MR-HMD play a crucial role in twofold. *On the human operator side*, natural language-based task prompt instructions are given and collected from operators wearing MR-HMD at the initial phase of the HRC task. These prompts are then fed into LLMs for adaptive task planning, aligning with the MR-HMD's vision perception module to extract scene information and specify HRC task settings in detail. *On the mobile manipulator side*, the human operator utilizes the MR-HMD to extract data streams that aid in the accomplishment of HRC tasks. These data streams consist of the first-person view RGB scene image sequences stream and the semantic vector stream. The semantic vector stream includes information about

* Corresponding author.
E-mail address: lihui.wang@iip.kth.se (L. Wang).

https://doi.org/10.1016/j.cirp.2024.04.003

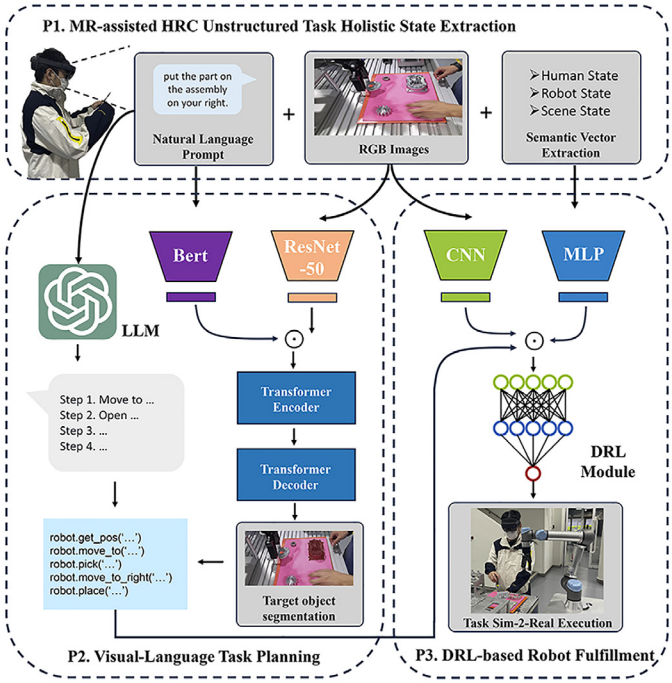

Fig. 1. System framework of the proposed human vision-language-guided DRL-enabled approach for unstructured HRC task fulfilment via MR-HMD.

pose and distance information among humans, robots, and goals detected with the device's own-built spatial functions, sensors, and detectors. Compared to conventional vision system settings, MR-HMDs are portable and mobile perception tools, that can perceive the scene flexibly on demand. Also, it brings the information of human and spatial information by nature without much extra computation cost. These dual data streams are processed and fed into a DRL-based whole-body robotic motion planning module. Consequently, it facilitates the optimization of the synergistic relationship between humans and robots during task execution, fostering a safe and efficient HRC in the unstructured environment.

2.1. Human vision-language-guided task planning

The ability to understand natural form of human vision-language commands and deduce a feasible robotic action plan is of tremendous importance for on-site HRC. Our preliminary works have introduced spatial-temporal neural networks [3] and multi-granularity scene segmentation strategy [4], to achieve HRC scene understanding. However, unlike explicit language communications, they rely solely on visual information, which can exhibit certain vagueness. Recent research endeavours, such as Venkatesh et al. [5] and Valente et al. [6] have leveraged human language cues to complement vision data for robotic task fulfilment. However, the language comprehension ability is still quite limited by the rather small-scale Long Short-Term Memory (LSTM) model. To address those issues, a vision-language-guided adaptive HRC task planning approach is introduced in Fig. 2,

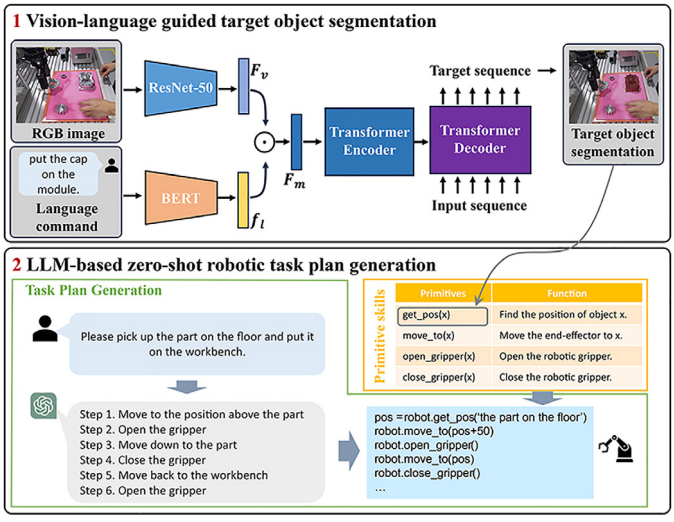

Fig. 2. Vision-language model-based adaptive HRC task planning approach.

which consists of two parts: 1) vision-language-guided target object segmentation, and 2) LLM-based zero-shot robotic task plan generation.

Vision-language-guided target object segmentation. It aims to segment a certain object from the visual observation based on the specification underlying the language command. Concretely, the input to the model consists of an RGB image of the HRC scene and a natural language command. The RGB image will first be processed by a visual encoder, such as a ResNet-50 backbone, owing to its wide adoption and great generalization ability, of which the result is the extracted visual feature $F_v \in \mathbb{R}^{(H\times W)\times C}$. H, W, C are the height, width, channels of the feature map, respectively, and $\mathbb{R}$ denotes that the elements are real-valued. The language data will be exploited by a pretrained language encoder BERT and transformed into a language feature $f_l \in \mathbb{R}^C$. The vision and language features are then fused into the multi-modal feature $F_m \in \mathbb{R}^{(H\times W)\times C}$ by Hadamard product. A standard Transformer [7] encoder is then adopted to process F_m. Then a Transformer decoder predicts the coordinates of the target object contour points in an auto-regressive manner, with a multi-layered perceptron and a final softmax function. The coordinate sequence can be further easily converted to the final segmentation mask via basic image processing. Sine positional encoding is added to the input sequence for permutation invariance. The training loss function is formulated as a cross-entropy loss over the sequence:

$$L = -\sum_{i=1}^{2N} w_i log P(\hat{y}_l | F_m, y_{1:i-1}), \tag{1}$$

where w is the per-token weight, y and $\hat{y}$ are the input and target sequences associated with F_m, and P stands for the estimated probability measuring the similarity between the estimation and ground truth. The primary advantage of this formulation over binary mask prediction works is that one can transform the 2D mask prediction task into a point sequence (mask contour point coordinates) prediction task. This is more aligned with sequential prediction nature of Transformer architectures. The target object segmentation model serves as the fundamental implementation of the *get_pos(x)* primitive skill, which is the most crucial functionality for further adaptive task planning.

LLM-based zero-shot robotic task plan generation. To achieve a zero-shot task planning, an LLM-based task plan generation approach is shown in Fig. 2, with an example of a human-guided robot pick-and-place task. Based on the derived segmentation information from the previous step, the LLM makes a parse of the human instructions and performs task planning. It divides the task into six steps corresponding to the order of movements as well as the locations indicated by the instruction. Then, code generation LLM is utilized to complete the robotic task by calling primitives according to the task steps.

Specifically, Generative Pre-trained Transformer (GPT-3.5) is leveraged to interpret natural language instructions and decompose them into subgoals for task planning. Additionally, we employ the Codex code generation model, which is pretrained on billions of code lines from GitHub, to synthesize executable Python robot code. This enables accurate invocation and composition of parameterized action primitives and execution of arithmetic operations when needed. The robot code is expressive of function or logic structures, and allows for parameterized Application Programming Interface (API) calls, such as *robot.move_to(object)*. Some other primitive robotic skills are shown in Fig. 2 as well. With the primitives in place, the model can receive new instructions and automatically combine API calls to generate new robot codes. Through the adaptation of LLMs and the integration of primitive skills, the approach demonstrates great potential for adaptive HRC task planning. The sequential primitive tasks will be further transmitted to the DRL module for the concrete planning and control of the mobile manipulator for task fulfilment.

2.2. DRL-based mobile manipulator task completion

With the planned tasks and segmented scene information from the previous section, the mobile manipulator is employed to fulfil the specific manufacturing tasks. Compared to fixed-base robots, the mobile manipulator exhibits significant potential for broader applications in HRC, owing to its extended range of movements and high flexibility in completing unstructured manufacturing tasks [8]. Constrained by factors such as control complexity, task execution safety, and programming costs, the predominant approach to mobile manipulation in HRC involves the sequential execution of movements of the mobile base and the robotic arm. However, such simplification compromises efficiency, impacting both control performance and productivity [9]. Inspired by [10], this research further adopted DRL in generating an end-to-end whole-body motion planning policy, facilitating the completion of HRC tasks in unstructured manufacturing scenarios via MR-HMD.

Problem formulation: DRL is an optimization approach that enables agents to self-optimize via autonomously interacting with the environment [11]. In this work, the mobile manipulation is formulated as Markov Decision Processes to generate the control policy $\pi(a_t|s_t)$ and optimized by DRL of generating action a_t

$\in A$ regarding state $s_t \in S$ to gain the largest cumulated reward (i.e. better performance) $E_{\tau\sim p_\theta(\tau)}\left[\sum_{t=0}^{T}\gamma^t r_t\right]$. Under this circumstance, with the previous assigned task primitive settings, the DRL module collects both the first-person/third-person view RGB image flow and the semantic state vectors via MR-HMD to predicts the actions a_t for the robotic moving base and robotic upper arm to fulfil the task primitives. The workflow of the proposed DRL-based planning policy is shown in Fig. 3. In which, the policy is trained by the Proximal Policy Optimization (PPO) algorithm [12]. PPO is widely recognized in the field of robotics [13] and it utilizes an actor-critic framework with an update clip function and importance sampling, to ensure efficiency and promote stable learning performance. The loss function of PPO is shown in Eq. (2):

$$L^{clip}(\theta) = E_\tau\left[\sum_{t=0}^{T} min\left(\rho_t(\theta)\hat{A}_t, clip(\rho_t(\theta), 1-\varepsilon, 1+\varepsilon)\right)\hat{A}_t\right] \quad (2)$$

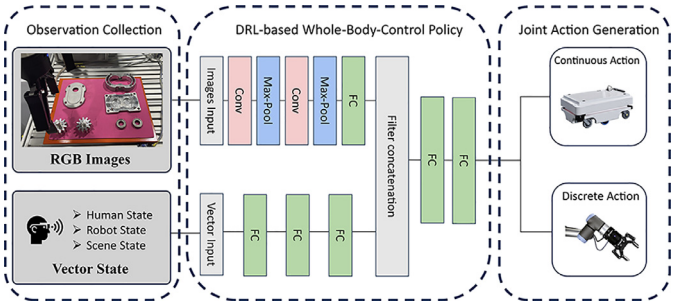

Fig. 3. Workflow of the proposed DRL-based mobile-manipulator whole-body motion planning approach.

The clip function is applied element-wise to each component of the ratio $\rho_t(\theta) = \frac{\pi_{\theta_{update}}(a_t|s_t)}{\pi_{\theta_{old}}(a_t|s_t)}$. If the ratio falls outside the range $[1-\epsilon, 1+\epsilon]$, it is clipped to the nearest bound. ϵ is a hyperparameter that controls the range within the ratio $\rho_t(\theta)$ clipped. This effectively limits the extent to which the policy can change at each update step. $\hat{A}_t$ is the advantage function that measures how much better an action is, compared to the average action at that state. Hereby, PPO achieves a balance between exploration and exploitation, preventing large policy changes while still allowing an effective learning and adaptation process.

In the context of DRL algorithms-based mobile manipulator control, the algorithm settings pertaining to the *state, action*, and *reward spaces* play a pivotal role. Unlike fixed-base robot control, mobile robot manipulation introduces additional complexities of collaboration between the base and actuator that requires careful consideration before its practical implementation. To ease the process, detailed algorithm settings are outlined below:

Observation state (S) serves as a representation of the human-robot working scene and its associated conditions. Unlike conventional observation approaches, the MR-HMD not only captures visual pixel signals as inputs for control policy, but also provides rich semantic vectors such as robots, human operators, and environment relevant information. The mixed state representation could significantly contribute to the learning process, and further improve the efficiency and performance in control policy for HRC activities. Meanwhile, the internal sensing system of the mobile manipulator is only adopted for emergency safety assurance purposes. Followed by the proposed approach, the MR-HMD collected state representation consists of two parts: 1) the visual input part S_i, which is from the MR-HMD first-person view camera and an external third-person view fixed camera, and 2) the state vector detected by MR-HMD devices S_{vect}, which consists of the *robot* (e.g., joint states, end effector), *human* (e.g., hand, body, head), *scene* (e.g., layout, obstacles) and *relative properties* (e.g., object-relevant distance information). Moreover, especially in an unstructured manufacturing scene, the temporal information, including obstacles and robot trajectories are important references for control policy generation. Thus, the temporal information of the past 2 frames' state vectors are concatenated and stored in the buffer to fit the policy learning with MR-HMD, *i.e.*, $\mathbf{S_t} = [[s_{img_t}, \cdots, s_{img_{t-2}}], [s_{vect_t}, \cdots, s_{vect_{t-2}}]]$.

Action space (A) is a joint space of discrete and continuous actions $\mathbf{A} = [a_{disc}, a_{conti}]$ for the whole-body motion control to handle the mobile manipulator geometry purpose. The discrete part maps to the robotic arm's end effector movement, in which the inverse kinematics solver is utilized to transform the robot's joint space into the Cartesian coordinates of fixed-oriented $a_{disc} = (\Delta x, \Delta y, \Delta z)$. The continuous one a_{conti} is the mobile robot base's two-dimensional actions for relative position control, including one dimension for forward/backward control and the other dimension to control the orientation of the mobile robot base, where $a_{conti} = (\Delta x, \Delta\theta)$. With such a mixed action space, the DRL algorithm outputs the action of robotic arm and mobile base to collaboratively explore the feasible trajectories.

Reward space (R) incorporates multiple criteria to reflect task performance. In DRL-based robot manipulations, the most intuitive approach to formulating reward configuration is solely based on the task target. However, due to sparse feedback, this approach results in a large search space. To accelerate policy convergence and improving performance, additional constraints are introduced in the form of reward terms. Hereby, the reward signals can be decomposed into discrete task goal state, safety, and a continuous distance-based task progress indicator. It combines multiple tolerance settings, like success rate (e.g., target reaching deviation $\leq$ 5 mm), safety (e.g., human-robot distance $\geq$ 50 mm), time tolerance ($\leq$ 40 s), and task progress (e.g., robot-target/end effector-target distance), denoted as $\boldsymbol{R} = (r_{task}, r_{safe}, r_{progress})$.

3. Case study

To demonstrate the performance of our proposed approach in handling unstructured HRC manufacturing tasks, comparative experiments are conducted on HRC assembly tasks in the lab environment. The experimental setup includes visual sensors (e.g., Azure Kinect), a HoloLens2 MR-HMD, a GPU server (RTX 3080), and a mobile robot (UR5E + MIR100), as shown in Fig. 4.

Fig. 4. Demonstrative HRC unstructured scenes in simulation and practice.

3.1. Vision-language reasoning for HRC task planning

To demonstrate the performance of the proposed vision-language-guided HRC task planning approach, the target object segmentation model is first evaluated on a dataset collected in our scenario. The dataset contains 463 pairs of image-text data with 370 for training and others for testing. Each pair of data consists of an RGB image and a reference expression text indicating a target object in the associated image. The input image size is 640 × 640, extracting feature size H and W at 1/32 the original size. The transformer has a 256-dimensional hidden feature, 6 encoders, and 3 decoders. The Adam optimizer with an initial learning rate of 5e-4 and batch size of 32 are used to train the model. The evaluation metric we adopted for the target object segmentation model is *mean Intersection over Union* (mIoU), which is widely used in segmentation tasks to measure the segmentation accuracy. Comparative experimental results are shown in Table 1, which demonstrates an obvious improvement of our proposed method against previous approaches. Meanwhile, the LLM-based task planning strategy is evaluated in an empirical way by asking human experts to check the feasibility and correctness of the LLM-generated task plan. The evaluation is performed for three pre-defined tasks abstracted from the HRC working process: 1) Fetch a specific part from the storage area, 2) place a gear into the case, and 3) pick a case cover and put it onto the module. Each task will be repeated 20 times, and the corresponding success rate is calculated and listed in Table 2, which demonstrates the ability of the proposed method to facilitate HRC task fulfilment. In case of failure, human inspection can help prompt the LLM for an adjusted plan, guided by explanations of the prior plan's shortcomings.

Table 1
Experimental results of the target object segmentation model.

Method	Components	mIoU
Yu et al. [14]	ResNet-101; Bi-LSTM	65.36
Luo et al. [15]	DarkNet-53; GRU	72.76
Ours	ResNet-50; BERT; Transformer decoder	**77.89**

Table 2
Experimental results of the LLM-based task planning strategy.

Task 1	Task 2	Task 3	Average success rate
17/20	19/20	15/20	85 %

3.2. DRL experiments for mobile manipulator task fulfilment

To demonstrate the effectiveness of the proposed MR-HMD-based DRL approach, Table 3 shows the DRL training parameters of 1000 different targets in the simulated working scene, based on the real HRC task settings, and the success rates of different algorithm settings are presented in Table 4. Notably, our time-based approach performs best, whereas the purely visual one fails to complete any task. Moreover, the remaining results highlight the MR-HMD's ability to enhance the control policy for unstructured HRC tasks meeting the demands of scenarios.

Table 3
Training parameters of the proposed DRL-based approach.

DRL training parameters	Batch size	MLP hidden units	CNN hidden units	Learning rate	Training steps
Value	5000	512:512: 512	32:64: 64	0.0003	1M

Table 4
Experimental results of the MR-assisted DRL motion planning.

Algorithm settings	**Average success rate**	**Reward**	**Episode length**
Visual Vector	–	−11.76	62.49 Steps
MR-assisted Vector	54.4 %	−7.397	103.1 Steps
Visual + MR-assisted Vector	86.6 %	−3.814	95.13 Steps
Time-based Visual + MR-assisted Vector	90.2 %	−2.939	89.01 Steps

3.3. Experimental result discussions

From the HRC task planning results, one can observe: 1) a considerable improvement of the vision-language-guided object segmentation performance, and 2) the feasibility of leveraging LLM as the HRC task planner. However, practical issues of LLMs still require further investigation, such as computational and network latency, due to the reliance on cloud-run GPT models. Meanwhile, from the motion planning results based on add-on information stream by MR-HMD, two major advantages can be observed: 1) the robot learning process became more efficient and interpretable, and 2) the uncertainties of the control strategy have been largely reduced. However, as a trade-off, the proposed DRL method requires more prior expert knowledge and consumes higher computational resources. Additionally, to ensure the scalability and adaptability of the MR-HMD system in various manufacturing scenarios will result in a higher overall expense.

4. Conclusions and future work

This work proposes a vision-language-guided DRL-enabled task planning approach for unstructured HRC in manufacturing. The main scientific contributions of it include: 1) MR-HMD modelling as an effective tool for data collection, communication, and state representation in HRC task settings; 2) a vision-language-guided target object segmentation model to provide localization information for robotic action goals, along with an LLM-based robotic task planning module; and 3) an MR-assisted time-aware DRL-based whole-body motion planning policy for mobile robot manipulators to fulfil various unstructured manufacturing tasks. Their performance has been evaluated via comparative experimental results. In future, both multi-modality large models-based robot learning, and advanced human-in-the-loop learning mechanisms will be explored for more intuitive and effective HRC.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.cirp.2024.04.003.

References

[1] Zheng P, Li S, Fan J, Li C, Wang L (2023) A Collaborative Intelligence-Based Approach for Handling Human-Robot Collaboration Uncertainties. *CIRP Annals* 72(1):1–4.
[2] Zheng P, Li S, Xia L, Wang L, Nassehi A (2022) A Visual Reasoning-Based Approach for Mutual-Cognitive Human-Robot Collaboration. *CIRP Annals* 71(1):377–380.
[3] Li S, Zheng P, Fan J, Wang L (2021) Towards Proactive Human Robot Collaborative Assembly: A Multimodal Transfer Learning-Enabled Action Prediction Approach. *IEEE Transactions on Industrial Electronics* 69:8579–8588.
[4] Fan J, Zheng P, Lee CKM (2022) A Multi-Granularity Scene Segmentation Network for Human-Robot Collaboration Environment Perception. *IEEE International Conference on Intelligent Robots and Systems*, 2105–2110.
[5] Venkatesh SG, Biswas A, Upadrashta R, Srinivasan V, Talukdar P, et al. (2021) Spatial Reasoning from Natural Language Instructions for Robot Manipulation. *IEEE International Conference on Robotics and Automation* : 11196–11202.
[6] Valente A, Pavesi G, Zamboni M, Carpanzano E (2022) Deliberative Robotics—A Novel Interactive Control Framework Enhancing Human-Robot Collaboration. *CIRP Annals* 71(1):21–24.
[7] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, et al. (2017) Attention is All You Need. *31st Conference On Neural Information Processing Systems (NIPS2017)*, Long Beach, CA, USA.
[8] Sun C, Edrzej Orbik J, Devin C, Yang B, Gupta A, et al. (2022) Fully Autonomous Real-World Reinforcement Learning with Applications to Mobile Manipulation. *Conference on Robot Learning, PMLR*, 308–319.
[9] Li C, Zheng P, Yin Y, Pang YM, Huo S (2023) An AR-Assisted Deep Reinforcement Learning-Based Approach Towards Mutual-Cognitive Safe Human-Robot Interaction. *Robotics and Computer-Integrated Manufacturing* 80:102471.
[10] Li C, Zheng P, Yin Y, Wang B, Wang L (2023) Deep Reinforcement Learning in Smart Manufacturing: A Review and Prospects. *CIRP Journal of Manufacturing Science and Technology* 40:75–101.
[11] Sutton RS, Barto AG (2018) *Reinforcement Learning: An Introduction*, MIT Press.
[12] Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O., 2017, Proximal Policy Optimization Algorithms, *arXiv preprint arXiv:1707.06347*.
[13] Zhou C, Huang B, Fränti P (2022) A Review of Motion Planning Algorithms for Intelligent Robots. *Journal of Intelligent Manufacturing* 33(2):387–424.
[14] Yu L, Lin Z, Shen X, Yang J, Lu X, et al. (2018) MAttNet: Modular Attention Network for Referring Expression Comprehension. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1307–1315.
[15] Luo G, Zhou Y, Sun X, Cao L, Wu C, et al. (2020) Multi-Task Collaborative Network for Joint Referring Expression Comprehension and Segmentation. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1034–10043.