Advanced Engineering Informatics 62 (2024) 102572

Contents lists available at ScienceDirect

Advanced Engineering Informatics

journal homepage: www.elsevier.com/locate/aei

Full length article

Embodied AI for dexterity-capable construction Robots: DEXBOT framework

Hengxu You [a], Tianyu Zhou [a], Qi Zhu [b,*], Yang Ye [a], Eric Jing Du [c,*]

[a] Engineering School of Sustainable Infrastructure & Environment, University of Florida, 1949 Stadium Road 454A Weil Hall, Gainesville, FL 32611, USA
[b] National Institute of Standards and Technology, Boulder, CO 08305, USA
[c] Engineering School of Sustainable Infrastructure & Environment, University of Florida, 1949 Stadium Road 460F Weil Hall, Gainesville, FL 32611, USA

ARTICLE INFO

Keywords:
Construction robots
Embodied AI
Dexterity

ABSTRACT

The construction industry is integrating robots into critical tasks at an accelerated pace, with the aim of enhancing efficiency, safety, and productivity. However, construction tasks requiring dexterity remain a challenge due to the need for precise movements, accurate perception, real-time decision-making, and a comprehensive understanding of the environment. To address these challenges, the introduction of embodied artificial intelligence (AI) represents a significant shift in robotic capabilities to enhance their alignment with the broader spectrum of construction settings. Rooted in cognitive science, embodied AI emphasizes the integration of an agent's physical form into its computational intelligence processes. It resembles how humans develop motor skills by interacting with physical world. This paper introduces DEXBOT, an exploratory framework for designing construction robots capable of high dexterity using embodied AI principles that mimics human strategies in complex tasks. The framework outlines six key perspectives for solving high-dexterity tasks with embodied AI: scene understanding, localization and motion planning, position-based control, force-based control, sequence planning, and correction decision-making. By presenting preliminary test cases for each perspective, the paper emphasizes the role of embodied AI in advancing dexterity level of construction robots. The DEXBOT framework is expected to encourage interdisciplinary collaboration of designing capable construction robots in the future.

1. Introduction

The construction industry has witnessed an increasing adoption of robots in a spectrum of tasks, with the goal of enhancing the efficiency, safety, and productivity of construction operations [1], such as the bricklaying robot [2], robotic welding [3], and UAV/UGV for inspection [4]. Despite the advancements, the need for sophisticated collaborative robots becomes apparent in high-dexterity construction tasks i.e., tasks requiring fine motor skills, precise manipulation, and adaptive responses to unpredictable environments, which continue to pose considerable challenges for robotic applications, such as pipefitting, instrument installation, and electrical wiring [5–7]. These tasks challenge conventional robot control methods by not only demanding technical precision (accurate perception, real-time decision-making, and semantic understanding [8]) but also an advanced understanding of human behaviors and needs. The dynamic, uncertain nature of construction sites, with varying materials, unforeseen obstacles, and uncertain human actions [9], further highlights the importance of developing robotic systems that can effectively interact and collaborate with human workers, adapting in real time to their actions and decisions.

This paper proposes that one of the enablers for robotics to effectively execute dexterous construction tasks is integrating embodied artificial intelligence (AI) in the design of robots. Embodied AI represents a branch of AI methods that focus on modeling the complex interactions between AI and the physical environment to enable a more robust adaptation to the environment [10]. Rooted in the process of humans developing motor skills, embodiment originally refers to the role of the body of an agent (e.g., humans) in shaping intelligence, which involves sensory perception, motor control, and adaptive learning from physical interactions and experiences [11]. Embodied AI emphasizes real-time sensory data incorporation, enabling agents to respond more effectively to dynamic, unstructured environments and human presence. [10]. This approach can enables robots to interact with the environment and human workers in a more sophisticated and nuanced way, facilitating more effective reactions in dynamic and unstructured

* Corresponding authors.
E-mail addresses: you.h@ufl.edu (H. You), zhoutianyu@ufl.edu (T. Zhou), qi.zhu@nist.gov (Q. Zhu), ye.yang@ufl.edu (Y. Ye), eric.du@essie.ufl.edu (E.J. Du).

https://doi.org/10.1016/j.aei.2024.102572
Received 11 January 2024; Received in revised form 22 March 2024; Accepted 25 April 2024
Available online 2 May 2024

environments [12]. Recent advances in embodied AI hold great potential for significantly improving the capabilities of construction robots to facilitate complex construction tasks.

Despite the potential of embodied AI for enabling dexterous construction robots, a systematic framework that can guide the high-level design and integration processes seems missing. The significance of such a framework is twofold. One the one hand, a framework for robotic embodied AI can establish design standards for robot sensing, motor control, and learning mechanisms, essential for high-dexterity tasks. On the other hand, a systematic framework can encourage collaboration, accelerate research progress, and ensure that advances in embodied AI can be effectively and safely transferred to real-world construction tasks [13]. This is particularly crucial for enhancing the robotic applications in construction, where the need for intuitive, responsive, and safe interactions between humans and robots is paramount. The lack of such a framework may attributed to challenges in bridging the gap between existing robotics and AI methods and their challenges in the diverse and dynamic construction environment. First, the field of embodied AI is still in its infancy, and researchers are still exploring the best approaches for integrating AI into any physical system including construction robots [11]. Second, the construction tasks are complex and multifaceted, presenting a significant challenge in developing a universal methodology framework that can be applied across different types of construction projects [14]. The dynamic nature of construction environments and workers further aggravates this challenge, as robots need to adapt and respond to changing conditions in real-time [15]. Consequently, studies in construction automation tend to focus on very specific applications and tasks. Finally, the integration of embodied AI in dexterity-capable construction robots requires an interdisciplinary research approach, as it involves the convergence of AI and robotics. However, these are traditionally separate fields, and therefore the development of a systematic methodology framework for integrating embodied AI in construction robots requires collaboration and knowledge-sharing between experts in multiple disciplines.

To address this gap, this paper proposes an enhancement through the DEXBOT framework based on the existing foundation. This enhancement emphasizes the infusion of embodied AI principles into established frameworks for the design of dexterity-capable construction robots. Rather than introducing an entirely new paradigm, DEXBOT is designed to augment existing models with capabilities that significantly elevate the potential for sophisticated human-robot interactions, adaptability, and precision in unstructured environments. It meticulously details the roles embodied AI can play in various aspects of construction robotics, including scene understanding, localization and motion planning, position-based and force-based control, as well as sequence planning and the crucial decision-making processes of correction, rework, or discard. These steps serve as building blocks for the construction of capable robots that are pivotal for robots to effectively work alongside humans in unstructured environments. For each step, we will present preliminary test cases to showcase the unique role of embodied AI. By demonstrating the potential benefits of embodied AI for construction robotics, this paper aims to encourage further research and development in this area, ultimately leading to more capable and efficient construction robots in performing high-dexterity tasks.

2. Literature review

2.1. HRI in construction

The construction industry is witnessing a transformative shift towards HRI, moving from traditional, isolated human and robot roles to a collaborative, integrated approach. This shift, driven by the industry's growing emphasis on efficiency, safety, and productivity, is redefining the landscape of construction operations [16]. The evolution of HRI in construction reflects both technological and cultural shifts. Robots are increasingly seen as collaborative partners rather than mere tools, a change driven by the need for enhanced safety in hazardous construction environments [17]. Robots in roles like structural assembly or material handling alleviate the risk to human workers, taking on tasks that are either too dangerous or physically demanding [18]. The advancements facilitating this integration are multifaceted. Innovations in sensor technology allow robots to have a heightened awareness of their human counterparts, leading to safer and more efficient collaboration [19].

Furthermore, developments in AI and machine learning equip robots with the ability to predict human actions, thereby enhancing collaborative efforts [20]. Deep Reinforcement Learning (DRL) and Deep Imitation Learning [21] emerge as key methodologies, with DRL optimizing decision-making through trial and error and DIL enhancing robots' ability to learn from human actions directly [22]. This shift underscores a strategic move towards integrating human expertise into robotic systems, especially in complex construction tasks requiring high levels of dexterity [23]. DIL methods, such as behavioral cloning and generative adversarial imitation learning, have proven particularly effective in narrowing the gap between robotic capabilities and human skill, by enabling robots to replicate expert human behaviors quickly [24]. This integration not only accelerates the learning curve but also enriches robots' operational versatility, making them more adept at responding to the nuanced dynamics of construction sites [22]. By emphasizing human demonstrations—from visual cues to force interactions—as fundamental elements of robot training, we underscore the growing importance of HRI in the construction industry [25,26]. This approach not only streamlines the process of robot training but also ensures that robots can work alongside humans more efficiently and safely, embodying the nuanced demands of construction tasks through a harmonious blend of human insight and robotic precision.

Despite these advancements, HRIs in construction face several challenges. Communication barriers between humans and robots remain a significant hurdle. Developing an intuitive, user-friendly framework for human-robot interaction is crucial for maximizing the efficiency and safety of these collaborations [27]. Ensuring the safety of human workers in close proximity to robots is another critical area requiring ongoing research and development [28]. The advancements and challenges in HRI underscore the need for a comprehensive framework that not only enhances human-robot collaboration but also leverages the latest technological innovations [29]. Embodied AI bridges the gap between advanced robotic capabilities and intuitive, efficient HRIs by enabling robots to understand and respond to the physical world in a more human-like manner [11].

2.2. Construction robot and design framework

In recent years, the construction literature has shown a growing interest in robotic methods to support construction operations, aiming to enhance efficiency, safety, and productivity [30]. The applications of construction robotics, from material handling [31] and site preparation [14] to search and rescue [32], structural assembly and installation [33], painting and plastering [34] and robot-based inspection [35–37], have not only improved operational efficiency but also facilitated novel Human-Robot Interactions (HRIs) in construction scenarios. These advanced applications demonstrate the potential for collaborative interactions between humans and machines, ensuring safety and quality control across various construction tasks.

Innovations in mechanical designs, dynamic control methods, and human-robot collaboration have been specifically tailored to meet the unique challenges and requirements of construction tasks. The exploration of mobile manipulators for construction waste management [38] and UAV-UGV collaboration for site inspection [39–42] highlights the diversity in mechanical designs aimed at task-specific functionality. Moreover, novel control methods and efforts to improve robots' environmental and scene understanding capabilities [43–45] reflect the push towards increasing adaptability and precision within complex work environments.

Recent research also underscores the active roles of human agents in robotics, with mixed reality for robot teleoperation [46] and the use of haptic controllers for intuitive interactions [47–49] being explored. However, meeting the dexterity and adaptivity requirements of construction tasks remains challenging. Robots capable of performing a range of simple manipulation actions encounter difficulties in contact-rich interactions [15], highlighting the need for more effective force-based controls and enhanced perception capabilities. The increasing demand for a synergistic blend of human expertise and robotic efficiency [50,51] underscores the imperative for advanced HRIs in construction robotics.

Existing work in construction robotics has proposed systematic frameworks addressing these challenges through innovations in design and integration processes [52–56]. Yet, the inclusion of embodied AI elements is increasingly recognized as essential to further enhance robot dexterity and adaptability in unstructured environments. Studies have emphasized embodied AI's contribution to scene understanding, localization, motion planning, and intuitive human-robot collaboration [57–59], bridging the gap between robotic capabilities and the nuanced demands of construction tasks. Furthermore, advancements in embodied AI have led to the exploration of force-based control methods and sequence planning tailored for construction robotics [57], enabling robots to engage in more sophisticated interactions and decision-making processes. The development of frameworks that integrate embodied AI for high-dexterity tasks has become a focal point for recent research [60], demonstrating its potential to significantly improve the capabilities of construction robots [61,62].

In summary, while considerable progress has been made in construction robotics, the integration of embodied AI elements into existing frameworks emerges as a pivotal step towards overcoming challenges related to dexterity and adaptability. This approach promises to revolutionize construction robotics, enabling robots to effectively work alongside humans in complex and dynamic environments, thus addressing the evolving challenges of modern construction sites.

2.3. Embodied AI for robotics

The concept of embodied AI stems from the paradigm of embodied cognition, which argues that cognitive processes are influenced by the physical attributes of the body [63]. In robotics, embodied AI entails the physical instantiation of an AI system in a robotic body that allows it to interact with and learn from its environment [12]. The argument is that an AI should not merely act as a separate entity (such as a decision unit based on pre-fed training data); instead, it should be integrated with the body to leverage sensory and motor systems that interact dynamically with the environment in real-time [64]. By incorporating embodied AI into robotics, robots can be equipped with advanced perception, planning, and control mechanisms, allowing them to perform complex tasks in real-world environments [65]. Moreover, the incorporation of AI-interpreted human intention emerges as a critical complement to these capabilities [66]. By understanding and anticipating human intentions, embodied AI can facilitate more intuitive and seamless human-robot interactions, effectively bridging the gap between human operators and robotic systems [67]. This nuanced interpretation of human intent allows robots to adjust their actions proactively [68], aligning with the operator's goals and preferences [69], thereby optimizing collaboration and efficiency in tasks requiring close human-robot cooperation. Researchers have been exploring the potential of embodied AI in robotics and have been developing novel techniques to enhance the capabilities of robots [70,71].

To be noted, the literature has noticed multiple challenges in realizing a full-span embodied intelligent robot. One of these challenges relates to robot perception and scene understanding abilities, which are critical for robots to interact effectively with their surroundings [72]. Various methods have been proposed to enhance robot perception, such as deep learning techniques for object recognition [73], semantic segmentation for scene understanding [74], and reinforcement learning for human-inspired perception designs [75]. These advances have enabled robots to better perceive and interpret their environment, paving the way for more complex interactions and tasks. Another critical component of embodied AI is the development of efficient and adaptive control mechanisms for robots. Traditional controllers rely mainly on position-based controls, where the goal is to achieve the desired positions [76]. Recently, the robotics literature has highlighted the importance of exploring better force-based or touch-based control methods, where robots can react to physical forces in a more direct way [77–79]. In a recent review study [65], researchers found that the ability of robots to sense and model force and contact information through tactile or other force sensors would be critical to realize embodied intelligence in robotics. Finally, researchers also noted that the fulfillment of embodied AI would require intelligent agents to make spontaneous sequence decisions based on the real-time feeding of physical features of the surrounding environment. For example, Batra, et al. [78] found that a challenge to be resolved for embodied intelligent robots should be the ability to sense, understand and model the physical characteristics of the scene and objects, and thus make an arrangement decision without human intervention.

In summary, the literature on embodied AI for robotics has seen considerable progress in recent years, with advances in perception, planning, and control mechanisms, as well as high-level sequence decision-making. There is an emergent need for a systematic framework that categorizes the roles of embodied AI for dexterity-capable robots in dynamic and unstructured environments such as construction, to benchmark the progress and challenges in the area. Addressing this gap in the literature can lead to the systematic development of more capable and efficient construction robots, which can revolutionize the construction industry by automating a wide range of dexterous assembly tasks that are currently only performed by humans.

Our framework, grounded in embodied AI, is tailored to meet the unique challenges of HRIs in construction. By focusing on key elements of dexterity and adaptability, it comprehensively integrates steps essential for embodied capabilities in robots. This includes advanced perception, adaptive control, and sophisticated learning mechanisms, ensuring effective and safe collaboration between robots and human workers. Our approach revolutionizes the role of robots in construction, moving towards a seamless integration of human expertise and robotic efficiency. This paradigm shift, enabled by embodied AI, equips construction robots with enhanced capabilities to perform complex tasks, transforming the construction industry with innovative assembly techniques and collaborative strategies.

3. Framework of dexterity-capable construction robot

3.1. Overall architecture

We focus on three major types of construction tasks: structural assembly, material handling, and quality inspection, which reflects construction's fundamental challenges, emphasizing the importance of precision, efficiency, and safety. These tasks, integral to construction success, illustrate our framework's versatility and impact, aligning closely with the industry's diverse requirements. In structural assembly, precision and safety are crucial due to the risk of structural failure [80]. Material handling's efficiency and adaptability are key due to diverse material characteristics, posing logistical and safety challenges [81]. For quality assessment, ensuring accuracy and consistency in defect detection is crucial to uphold construction standards and avoid rework [82]. To address the concerns of the above tasks, we propose the framework for using embodied AI to enable dexterity-capable robots for construction tasks, or the DEXBOT framework. Fig. 1 illustrates the framework of DEXBOT.

To ensure robots interpret environments and navigate efficiently, we

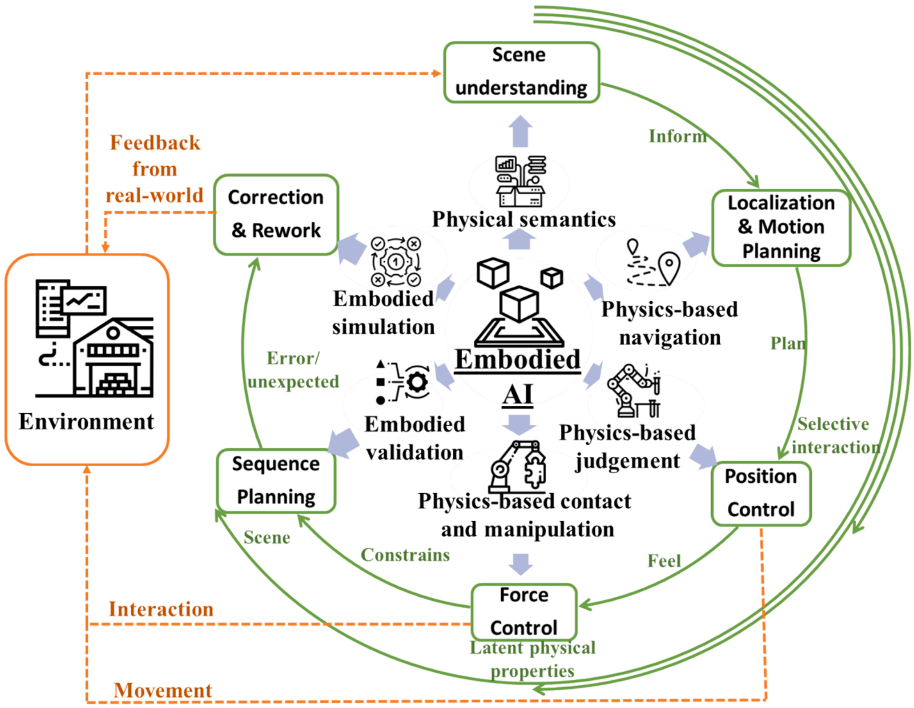

Fig. 1. Architecture of DEXBOT framework; critical steps rely on embodied AI that learns from the physical interactions with the environment.

propose the scene understanding and localization & motion planning to provide a foundation for precision in assembly, material handling and inspection in construction. The inclusion of scene understanding as a foundational step is motivated by the necessity for robots to interpret complex and dynamic environments accurately. This capability is central to embodied AI, which emphasizes the importance of sensory data in informing robotic perception and action [83]. For the scene understanding phase, embodied AI provides rich physical information and semantics for the robot to interpret the environment, identify individual objects, their dimensions, and positions, and more importantly, the physical properties and relationships among the objects.

The localization and motion planning are derived from the essential requirements for autonomous navigation and task execution in unstructured construction environments. Embodied AI posits that effective action is contingent upon an agent's understanding of its position within a space and its ability to plan movements that are coherent with the physical laws and environmental constraints [84]. In this step, the robot must accurately map its physical position within the workspace, identify the location of assembly components, and chart a clear, efficient trajectory to move from one point to another while avoiding any obstacles. These first two steps resemble how a human worker estimates physical properties of obstacles along the navigation path, not only including the kinematics but also the potential interaction with the environment, such as moving obstacles away and identifying potential hazards. Embodied AI provides the foundation for perceiving the physical properties of the environment.

To enable precise manipulation and interaction with various materials, we propose the position-based control and force-based control, which is key for assembling construction components accurately, handling materials safely and performing detailed inspections [85]. The distinction between position-based and force-based control methods reflects the nuanced requirements of construction tasks, which often demand a combination of precision and adaptability. The theoretical foundation for these control strategies is situated in the concept of embodied interaction, where the robot must not only execute predefined paths but also respond to real-time feedback and physical resistances encountered during construction task execution [86]. Position-based controls enable the robots to maneuver accurately in the task space, pick up objects, and align them precisely for assembly, while also adjusting to small deviations in real-time. As for a human worker, this step requires understandings of objects' physical properties, such as texture, weight, material, and geometry, for the optimal selection of objects and position control mauver strategies. Embodied AI could grant robot the same level of understanding. Force-based control is needed for the robot to sense the force interaction and apply the force correspondingly for assembly tasks such as screwing, pressing, or fitting parts together. Embodied imitation learning can be utilized to learn from human workers involving the delicate balance, applying enough force to accomplish the task without damaging the components or the robot itself.

Based on the scene understanding results and constraints of the force-based control, step 5, sequence decision, is needed for robots to determine the optimal order of actions. The inclusion of sequence planning acknowledges the complexity of construction tasks that require the execution of multiple, interdependent actions. Embodied AI, with its emphasis on adaptive behavior based on environmental feedback, provides a theoretical framework for understanding how robots can optimize construction task sequences to enhance efficiency and effectiveness in real-world scenarios [11]. This involves an understanding of dependencies between actions, the constraints, and the potential for time efficiencies in the order they are executed. Embodied AI can play a vital role in verifying algorithm-based sequence decisions by examining the physical constraints, such as the physical capabilities of assembly components and the possible interactions with the environment.

Finally, after the robot exerts the actions in the real world, the feedback is analyzed in step 6. This step addresses the critical need for robots to not only perform tasks but also to evaluate the outcomes and decide on subsequent actions. This capacity for self-assessment and decision-making is a hallmark of advanced embodied AI systems, drawing on theories of autonomous decision-making and machine learning to enable robots to learn from their actions and improve task performance over time [87,88]. Here, like an experienced human worker, the robot evaluates the results of its actions, compares them to the intended outcome, and decides if corrective actions are needed or if the task was completed successfully. If the construction task is successful, the robot can move on to the next task or, if necessary, iterate the current task with the revised parameters. Embodied AI can provide robots with the capability to learn from these feedback loops, enhancing their adaptive responses over time. These second two steps enhance all tasks by optimizing action sequences and adapting to feedback, crucial for streamlining assembly processes, ensuring efficient material management, and maintaining high-quality standards.

In summary, embodied AI serves as a cornerstone in uniting the six crucial stages of our framework. Initially, it processes sensory data for scene understanding, laying the groundwork for accurate localization and adaptive motion planning. This comprehensive environmental awareness enables the robot to execute actions with precision, using position-based and force-based control methods tailored to the task's specific requirements. Sequence planning then leverages this integrated data, allowing the robot to execute a series of interconnected actions aimed at achieving complex objectives. Crucially, embodied AI evaluates the outcomes of these actions, enabling the robot to make informed decisions about correction, rework, or alternative strategies as needed. This iterative cycle of action, evaluation, and adaptation—rooted in the robot's sensory and motor systems—illustrates how embodied AI not only interlinks these stages but also perpetuates continuous improvement and learning, thereby optimizing the robot's interaction with and adaptation to its operational environment.

We propose that the integration of embodied design principles is fundamental for realizing embodied intelligence in robotics for dexterous tasks. This novel approach creates a holistic, adaptive system capable of navigating complex physical environments, interpreting assembly goals, executing precise movements, applying appropriate force, efficiently deciding on task sequences, and learning from its actions. Each step of our framework, from scene understanding to decision-making, is interconnected and mutually enhanced by embodied design, emphasizing its vital role in enhancing robot dexterity and human-robot interaction in construction settings. The subsequent sections will delve into the technical details of each step, supported by case studies that illustrate the practical applications and benefits of our integrated approach.

3.2. Scene understanding

Developing a robust ability for scene understanding is the first step towards integrating embodied AI into construction robots. The construction environment is highly dynamic and unorganized, presenting unique challenges that demand effective perception and interpretation capabilities [89]. These challenges have been addressed through various existing methods that enhance a robot's scene understanding. Computer vision methods can help robots identify and understand the various objects and structures within the construction site via object recognition and semantic segmentation based on imagery data collected from sensors like LiDAR and RGB-D cameras [90]. These systems are also attuned to recognize and differentiate human workers, ensuring safe and intuitive human-robot collaboration. Semantic segmentation further aids in this by providing detailed contextual information about the environment, enhancing the robot's ability to interact appropriately with both the physical site and human workers [91]. These techniques can provide detailed information about the objects' positions, orientations, and semantic relationships, supporting more nuanced and safe interactions between robots and construction personnel [92].

While the recent advances in scene understanding have substantially promoted the quality of semantic outcomes, i.e., understanding the categories, identities, and contextual meanings of objects [93], we argue that granting robots the human-like ability to comprehend the physical properties of identified objects is also critical to enhancing their capabilities. Specifically, detected objects should be mapped with their physical categorizations, thus enriching the semantic understanding with a deeper, physical comprehension of the environment [94]. Such a mapping allows the robot to better understand the implications of interacting with these objects physically, or what we call, embodied sensing. Embodied AI offers tremendous potential for facilitating the development of embodied sensing which is the foundation for better maneuverability of construction robots [95].

Fig. 2 shows the workflow of how the proposed embodied sensing can enhance the scene understanding capabilities of construction robots. The case involves an intelligent agent (such as a construction robot or the AI that controls it) identifying stacked pipes along with an estimation of their key physical properties (such as weight and materials). Embodied sensing utilizes recognized scenes and objects (based on point cloud) with virtual objects (i.e., prefabs) in-game engine, and assigns physics properties and interactions based on the elemental models established by Universal Scene Description (USD) [96]. First, we utilized an adaptive LiDAR scanning method developed by You, et al. [9] to generate the augmented dense point cloud of the objects and environment. The scanned point cloud was fed to a density-based clustering for object identification. Then, we employed 3D point cloud detection algorithms, such as PointNet++ network [97] to detect the type of the segmented point cloud and estimate the pose. Afterward, a shared prefab library that contains objects in the working scene will be utilized. All objects collected in the library are provided with IDs, classes (scenery objects and dynamic objects), pose, quality, dimensions, and prefab models. Preparatory works include the collection of dimensional parameters of main objects and virtual object modeling. Then, raw point cloud data is replaced with corresponding physics prefabs. The object's key information (name, dimensions, pose) is subscribed from ROS. The name and dimensions of the identified object are used as the search key in the prefab library. Based on the mapping relationships between an object and the prefab, a virtual object with all physical properties in the library is generated via USD. A key decision point is what objects in the working scene should be replaced with virtual models for enhanced physics simulation at what time point. We define two classes for adaptive physics modeling, including scenery objects (such as environment and stationary structures), and dynamic objects (such as payload objects and other movable objects interacting with the robot). Scenery objects will be replaced immediately at the early phase of the process, while dynamic objects will be replaced based on their relevance to the task and the accumulation of motion and kinematics data. The last step is to call the physics USD schema to retrieve pre-established physics simulation data for expedited physics modeling and simulation. To simplify the physics simulation while still capturing representative physics processes in heavy rigging, the rigid body primer is recommended [98]. In our framework, a rigid body is described by its pose (position and orientation), as well as its mass distribution (center of mass position and an inertia tensor). The body also has a velocity (linear and angular vectors). Given the state, or the state history of the bodies at a specific time, we compute the updated state of the bodies a moment in time later, with the general desire being that the bodies' movement while constrained by the constraints obeys the laws of physics.

To illustrate the effectiveness of the proposed method, we tested our method on stacked object detection which is a common scene in construction site. Firstly, the robot would try to identify the location of each single pipe based on the point cloud detection results. Secondly, it would apply the PointNet++ detection network to classify the segmented single pipe and predict its label which is associated with the prefab in USD. Fig. 3 showed the segmentation results of the stacked pipe. Different colors corresponding to different pipes. Table 1 listed the 3D IoU results for each pipe and the overall evaluation. The visual results and the IoU accuracies showed that all the items could be well identified and segmented.

Additionally, we listed the classification results in Table 1 based on the segmented pipe to map to the prefab ID. Note that the primary goal was to ensure the detected pipe was accurately recognized as its specific prefab category, rather than broadly categorizing it among various prefabs. The emphasis was not on the overall classification accuracy of

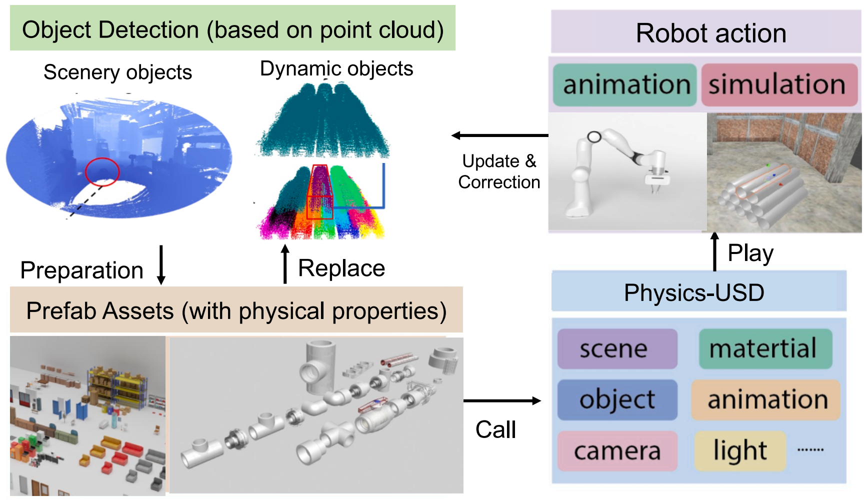

Fig. 2. Workflow of the embodied sensing for scene understanding including physical properties alignment.

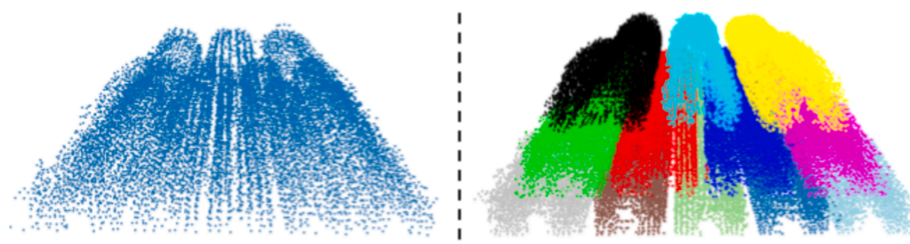

Fig. 3. Pipe segmentation results with the original point cloud (left) and the segmented result (right).

Table 1
3D IoU for each pipe and classification results.

Pipe Number	1	2	3	4	5	6	7	8	9	10
3D IoU (%)	90.2 %	94.6 %	91.3 %	93.4 %	89.6 %	91,7%	88.9 %	85.6 %	88.9 %	90.6 %
Classified Label	“pipe”	“pipe”	“pipe”	“pipe”	“pipe”	“pipe”	“pipe”	“pipe”	“pipe”	“pipe”

the detection network across multiple object types but on its ability to correctly map a segmented object to a particular type.

From the above case, we show that by developing a robust physics-modeling ability for scene understanding, embodied AI can enable construction robots to better comprehend their dynamic and unorganized surroundings. The method for incorporating physical properties into the objects of digital twin models should expedite the learning process of the AI system with precise models of real-world physics. This enhanced perception and interpretation capability is a critical step for integrating embodied AI in construction robotics.

3.3. Localization and motion planning

Localization and motion planning that suit well for dynamic and unorganized construction workplaces is another fundamental capability for dexterity-capable construction robots. It includes the need for both the navigation of the entire robotic platform and the manipulative motions to collaborate with human workers. Unlike robotic applications in well-controlled environments where objects, tools, and resources are placed at relatively fixed locations, construction robots often need to accurately locate a dynamic entity and move it along a desired path [99]. The need for effective localization and motion planning techniques arises from the confined spaces of most construction workplaces due to the unique geometries of built structures, irregular layouts, and the presence of numerous workers [100]. These factors make it difficult for robots to perform tasks that require precise object manipulation and transportation. To address these challenges, control-based or learning-based techniques have been developed and explored, such as Simultaneous Localization and Mapping (SLAM), advanced control methods, heuristic search algorithms, and reinforcement learning methods [101–103].

Despite the efficacy of the above methods, construction robotics faces unique challenges including uncertainties in the environment (e.g., unexpected obstacles or changes in the terrain), uncertainties in the robot’s sensory data (due to sensor noise or inaccuracies), uncertainties in the robot’s actions (e.g., due to control errors or mechanical failures), and especially the uncertainties related to human workers (e.g., unpredictable human behaviors) [54]. These uncertainties also make it difficult for robots to make accurate predictions about the consequences of their actions and to plan their movements effectively. We propose that embodied AI could offer potential solutions to the issues associated with uncertainties in construction environments. Embodied AI allows agents to interact within a simulated environment that accurately mirrors the complexity and unpredictability of real-world conditions [104]. Using embodied AI, robots can be trained in a simulated construction environment that closely mimics the real world, encompassing a variety of scenarios such as unexpected obstacles, changes in terrain, variable lighting conditions, and the presence of dust or debris. Virtual training allows robots to better adapt their understanding of the environment, update path planning, and control strategies more efficiently, even in the face of inherent uncertainties [105]. Moreover, unpredictable human behaviors can also be modeled in these simulators, enabling the robots to continuously learn and adapt their behavior to minimize the impact of these uncertainties on their performance. The simulated environment also offers the advantage of allowing robots to make mistakes and learn from them without causing any real-world injury or delay [106]. Robots trained in a diverse range of simulated scenarios could better navigate complex environments and perform complex tasks with high levels of autonomy and precision.

In the following test case, we show the comparison of path planning results using the traditional visual-based SLAM method and embodied reinforcement learning [98] method. As shown in Fig. 4, all objects are assigned physical properties that enable them to provide similar-to-real collision and force feedback when interacting with the mobile robot. Some objects are relatively static (e.g., walls) and some objects are movable (e.g., cardboard boxes). This environment simulates a close-to-real-life navigation scenario in which objects can be interacted with at different extents. For the visual-based method, the agent uses the LiDAR sensor to capture the spatial information and uses the point cloud visual SLAM method to build the map and find the path to the target (see Fig. 5). The robot is required to use the visual inputs to avoid collision with the obstacles and find the path to access the target. For the ERL group, the agent uses the same sensors to explore the environment, while being allowed to interact with objects. A proximal policy optimization-based reinforcement learning kernel is applied to both conditions to train the agent to find the target. Negative rewards are given for hard colliding with static items, and positive rewards are given for reaching the target. In addition, for every step the agent takes it would get a negative reward so that it would be forced to find the fastest path by either shifting the soft item or passing through the gap between the obstacles and trying to avoid collision with hard item at the same time.

Table 2 lists the collision counts with hard object, soft object and the completion steps of the two groups. We conducted 100 inferences for each group and calculated the averages and derivations. Our result indicates the advantages of ERL over classic vSLAM in planning the optimal path on this construction site. Specifically, with both methods, the agent made numerous collisions at the beginning exploration stage, but it gradually learned to avoid collision with hard items. The vSLAM relies on avoidance algorithm to update the path, while ERL learns the physical properties, especially the moveability of the objects, to decide more proactive actions to the obstacles. According to Table 2, the collision count with soft object of ERL was significantly higher than that of vSLAM, indicating that the ERL agent learns to interact with the moveable objects to clean a path. On average, the vSLAM agent took more than 2,000 action frames to reach the target while the ERL agent

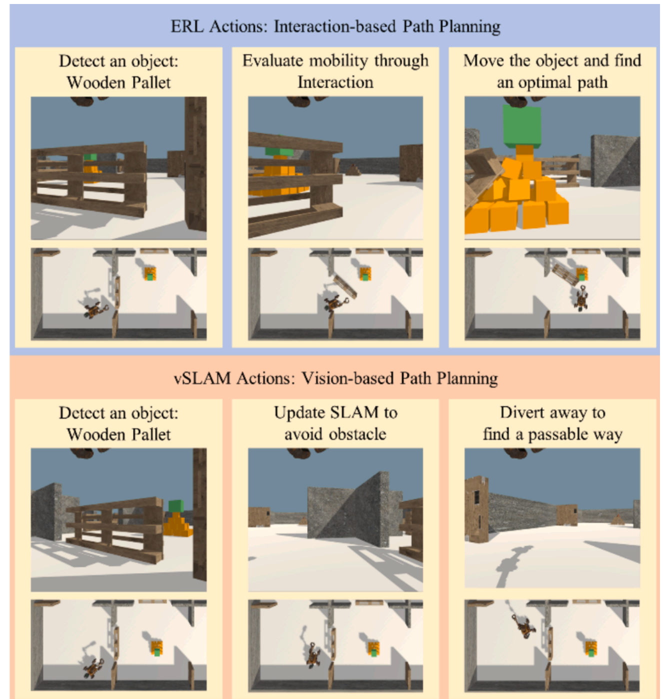

Fig. 5. Different Actions by vSLAM vs ERL.

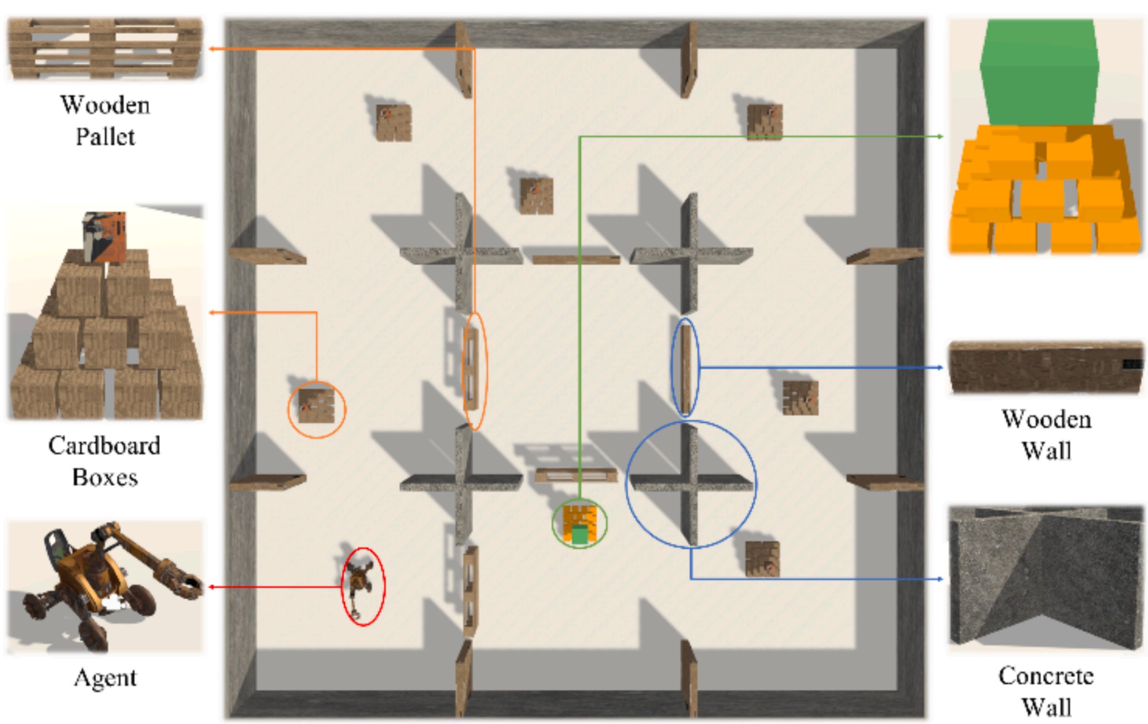

Fig. 4. Test environment of path planning.

Table 2
Inference results of two groups.

Groups	Collision Soft (number of frames)	Collision Hard (number of frames)	Completion Steps (number of frames)	Success rate
vSLAM	11	19	2134	95.6 %
ERL	192	21	475	98.6 %

took an average of 475 frames to reach the target. Also, the collision count with hard object of ERL was almost zero, meaning that the agent could identify which object was dangerous and couldn't be moved.

Without embodiment, the traditional vSLAM method could only find a sub-optimal path that avoided all obstacles. On the contrary, robot agents in the ERL group could learn to find the target through interacting with the surrounding environment. Besides, the agent developed the strategy to shift the soft item in the complex environment. The agent finally learned to move the pallet to create a passable way from its start point to the target within a short period of time. Therefore, compared with the traditional vSLAM-based method, the ERL method could handle the changeable environment and find the potential optimal solution through exploration.

3.4. Position-based control

Position-based control is another critical capability for dexterous construction robots, as it directly impacts their ability to perform complex object manipulation tasks with high precision and accuracy. Construction tasks such as bricklaying, pipefitting, and assembly often require intricate object manipulation, hence the ability to accurately pick and place objects, move objects, and align objects with their intended positions is of utmost importance [14]. Position-based controls can enable construction robots to effectively align objects in their operations. Position-based control focuses on regulating the position and orientation of the robot's end-effector, ensuring that it accurately follows a desired trajectory [107]. By leveraging the latest advancements in position-based control, construction robots can better manipulate objects and align them even in the presence of uncertainties and disturbances. Song, et al. [92] thoroughly examined the application areas of position-based controls for robotic manipulation and found that any application scenarios requiring motion compliance or delicate interaction should consider impedance position-based control as a potential solution. Khalil and Payeur [93] proposed a multisensory fusion method for improving the accuracy of position-based controls for manipulation tasks, even when the robot's model or the environment was not perfectly known. By improving these controls, construction robots become more adept at working in close proximity to human workers, ensuring precise and safe interactions during complex collaborative tasks.

Despite the advancements in position-based control methods, several challenges remain in achieving effective object alignment in dexterous construction robots. One such challenge is that there is an inherent variability of construction materials in sizes, shapes, and even physical properties. Different materials like concrete, steel, wood, or plastic may require different handling techniques and alignment strategies. For instance, aligning a steel beam for assembly requires different precision compared to aligning a plastic pipe for installation in terms of maximum speed, grabbing point, and safe overhead zones. Different materials can also be affected by environmental conditions in unique ways, such as how changing light conditions affect the reflection of materials, and later, affect the ability of the robot to correctly detect and grab the objects.

We propose that embodied AI provides a sound solution for refining position-based control that adapts to the complexity of construction scenes. Embodied AI in a virtual environment provides an opportunity for AI systems to gain a more nuanced understanding of control systems. The AI system can learn from thousands of virtual experiments, with precise reproduction of real-world environment conditions (e.g., texture, shapes, etc.), on how different parameters affect position-based control under various conditions [108]. For instance, an embodied AI system could sense the physical properties such as weight and texture and estimate the environmental interactions to adjust its position-based control strategy, such as speed and trajectory. Furthermore, embodied AI can enhance the feasibility of position-based control methods in manipulating deformable objects, which is considered nontrivial with traditional control methods, but is common in construction such as cable handling [109].

In the following test case, we show an embodied AI training architecture that uses a physics simulator with varying target object materials [110]. We aim to design and examine an embodied robot teleoperation system integrating a mixed reality simulator and a high-resolution haptic feedback system. Fig. 6 shows the workflow of the proposed structure. This simulator is built on a physical engine that can accurately simulate the physical properties of objects and the environment. The objective of the embodied AI is to control a robot arm to perform a pipe installation task efficiently. There are three types of pipes which are made of different materials and correspond to different weights and textures as illustrated in Table 3. Each pipe type had distinctive properties in terms of mass and friction. The distinct weight and texture require different position-control strategies such as gripping location, motion speed, and trajectories (e.g., the number of turns). For example, the PVC pipe, due to its lower mass, could be grasped with less pressure. However, if excessive pressure was applied, the PVC pipe could deform. The cast-iron pipe, on the other hand, was able to withstand a significant amount of pressure without deforming but required more force to pick up.

Fig. 7 showed the training setup of our experiments. We trained the robot arm in the simulated embodied environment to learn to apply proper grasping force and safely insert the pipe into the target outer pipes using reinforcement learning. To emphasis the importance of embodied learning, we designed two training group. For the control group, the observation of the agent only included the 3D location of the object and the target. The robot arm couldn't feel the physical properties. For the experimental group, the robot arm was provided the density, friction, pick-up force and deformation force plus the locations of object and target. To evaluate the performance of the trained agents, we calculated the success rates for the two agents. Either dropping the object or destroying the object with large grasping force would lead to failure. Fig. 8 showed an example of the trajectories of the two groups with the left figure standing for control and the right figure standing for experimental group. It is obvious that the agent with embodied information performed smoother and stabler trajectories compared with that of the non-embodied agent.

Additionally, we counted the completing steps, location errors and success rates for the two groups when handling different types of objects as shown in Table 4. For the control group, the completion time of different objects were almost the same. On the contrary, for the experiment group, the completion time of heavier objects (Cast-Iron) were larger while those of the lighter objects (PVC) were smaller. The similarity of control group and the difference of experiment group showed that embodied group could adjust the moving strategy according to the change of object's weight. Given that the distances of object shifting were almost the same, the robot agent developed a stable trajectory with slow speed for heavier object and a direct trajectory with fast speed for lighter object. Also, the overall radius errors of insertion (the shift between the object pipe center to target outer pipe center) for embodied group are lower than that of the non-embodied group. Consequently, the agent with embodied information has a significant higher success rate, indicating that the robot could leverage the additional embodied information to improve its performance.

This comparison showed that the robot agent could successfully determine different object properties by feeling the weight, inertia, and

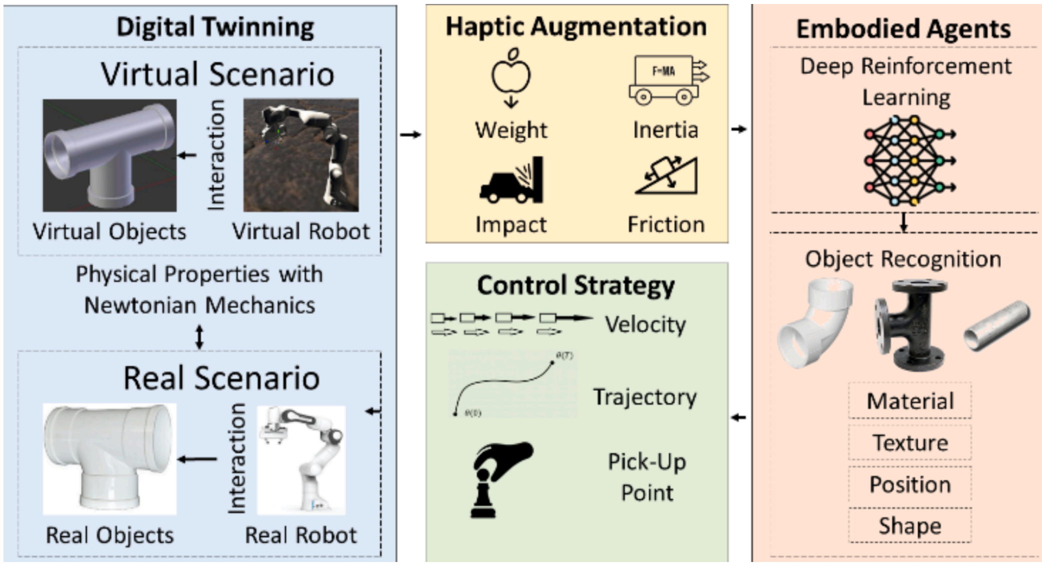

Fig. 6. Robot Teleoperation in a pipe operation task.

Table 3
Properties of different pipe materials.

Objects	Density	Color	Friction	Pick-up force	Deformation force
PVC pipe	1.4 g/cm^3	White rigid plastic	Medium	Small	Small
Aluminum pipe	2.7 g/cm^3	Bronze metallic	Small	Medium	Medium
Cast-iron pipe	7.3 g/cm^3	Dull black with a rough	Large	Large	Large

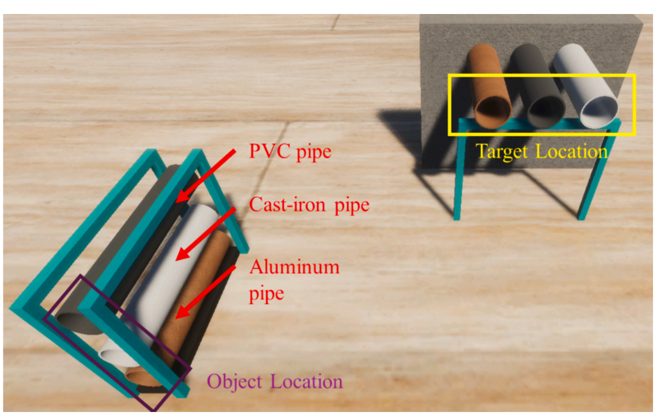

Fig. 7. Pipe operation training layout.

friction. Furthermore, it learned to adapt different position control strategies accordingly.

3.5. Force-based control

Another unique ability needed for dexterity-capable construction robots is the force-based control that enhances robots to perform manipulative tasks with varying physical properties [111]. This precise force application is especially important in scenarios where robots collaborate with human workers, as it enhances safety and task efficiency [112]. Given the importance of force control, there is a growing interest in developing methods that enable construction robots to exhibit versatile force control capabilities. These methods encompass a wide array of techniques, such as impedance control [113], force-torque control [114], and haptic feedback mechanisms [115]. These methods allow for flexible interaction with the environment, as the robot can adapt its movements based on the forces it encounters, making it highly suitable for handling delicate materials or interacting with other objects or people in a safe manner [116]. The literature has also examined hybrid force/position control, a method that controls both the position and the force exerted by the robot simultaneously. This is especially beneficial in tasks that require a specific force application along a defined trajectory, like gluing or painting [117]. Additionally, the integration of deep learning techniques with these control methods is gaining attention, which enables robots to learn and adapt their force control strategies based on their experiences, thereby enhancing their performance in various manipulative tasks [118]. These methods allow robots to adapt to physical interactions with both the environment and human collaborators, ensuring safe and effective joint task execution. Integrating such advanced force control capabilities in construction robots significantly contributes to a more harmonious and productive human-robot partnership in complex construction tasks.

A critical insight from our work with the DEXBOT framework is the recognition that ERL particularly when it integrates force information within a physics simulator, offers distinct advantages. Traditional reinforcement learning methodologies, while robust, often suffer from a lack of contextual understanding, particularly when nuanced physical interactions come into play. Incorporating force data into the learning process within a physics simulator offers the reinforcement learning agent a more comprehensive sensory palette [119]. This enriched data environment helps the agent to develop more nuanced policies that better account for real-world physical interactions. The agent, equipped with this added layer of sensory information, can simulate and predict outcomes with higher accuracy than when operating on visual or positional data alone [120].

The following case demonstrates our approach through a specific construction task – pipe inserting as shown in Fig. 9. The objective of the AI agent is to perform a dexterous pipe inserting task in an occluded operation space (i.e., accurate visual capture of the contact surface is infeasible to acquire) as shown in Fig. 9(a). We trained the AI agent with traditional reinforcement learning and embodied reinforcement learning. Only positional information from the visual sensor was provided to the traditional RL, while only force contact information was provided to the embodied RL. We trained 5 trials for each group. Fig. 9

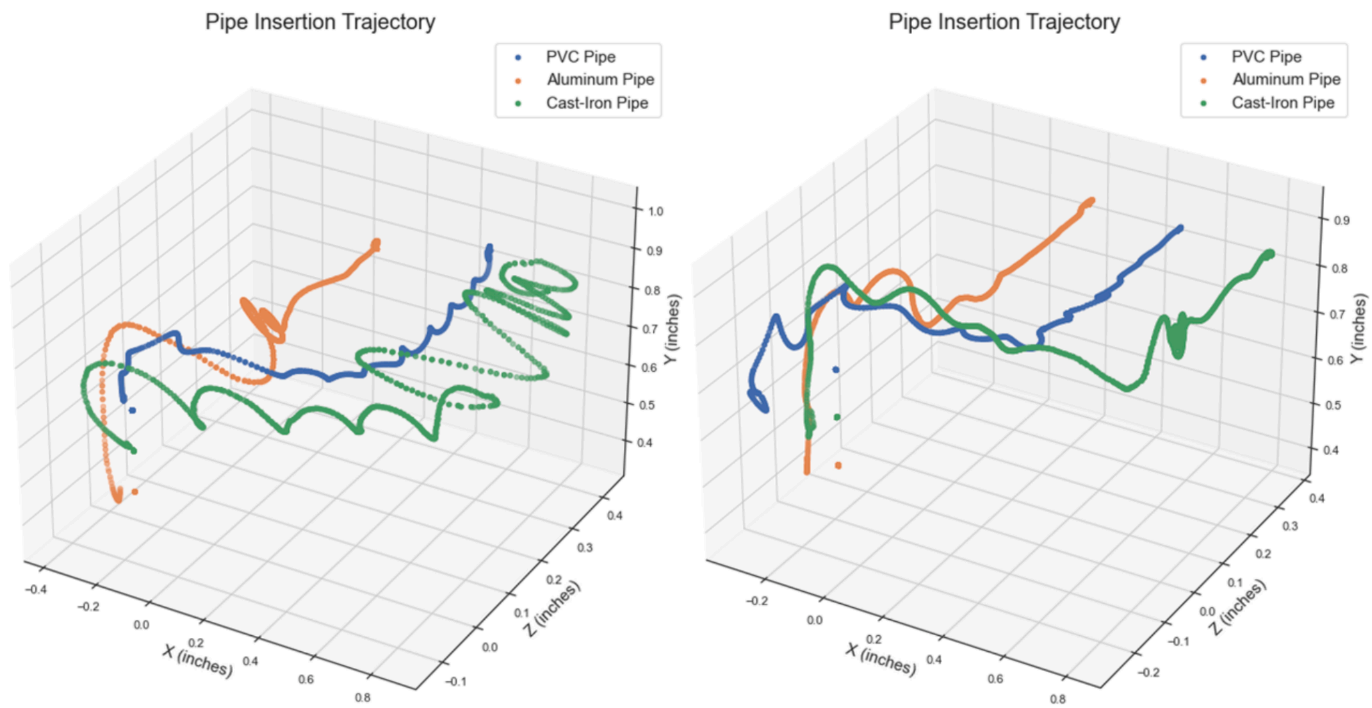

Fig. 8. Trajectories with embodied information (right) and without embodied information (left).

Table 4
Testing results of the two groups.

Groups	Pipe Material Types	Completion Time (s)	Radius Error (m)	Success Rate (%)
Embodied Group	PVC	16.57	0.0261	95.2 %
	Aluminum	17.21	0.0242	93.7 %
	Cast-Iron	16.23	0.0196	96.3 %
Non-Embodied Group	PVC	14.34	0.0375	80.6 %
	Aluminum	17.18	0.0212	74.2 %
	Cast-Iron	21.86	0.0561	69.7 %

(b) and Fig. 9(c) show the training and testing results, respectively. The y-axis in Fig. 9(c) denotes the steps to finish the task using the trained models. According to the comparison, we noticed that the time step for force-based group was largely smaller than that of the visual group, indicating that using force sensory data as embodied feedback could significantly shorten the task completion speed. Additionally, we listed the error and task success rate of the two groups in Table 5. The force group performed better than the visual group according to both metrics, which further verified the effectiveness of embodiment training.

Our experiments have consistently shown that agents trained through ERL consistently outperform their counterparts trained using traditional reinforcement learning methods. Not only do these agents achieve their objectives more effectively, but they also adapt more swiftly to unanticipated challenges or changes in their environment, underscoring the potential of embodied reinforcement learning as a transformative tool in robotic manipulation and beyond. Moreover, we also found that the learning speed and accuracy were significantly improved while imitating human demonstration data. This result provides a fertile ground for future human-robot collaboration study and embodied imitation learning.

3.6. *Sequence planning*

Sequence planning also plays a vital role in the performance of dexterity-capable construction robots. A construction task often consists of a variety of interdependent steps that must be executed in a specific

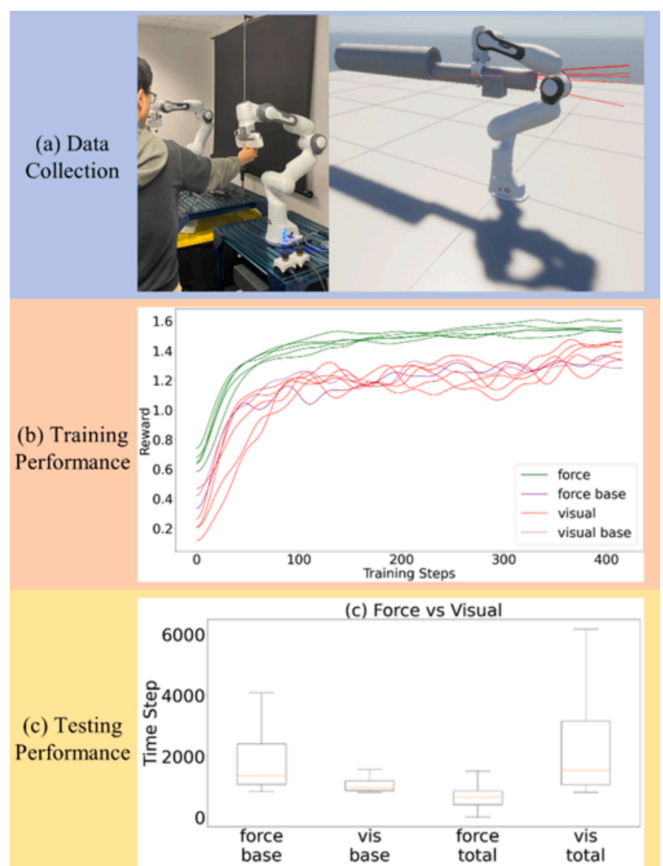

Fig. 9. Embodied AI for robot force control: an experiment comparing the training results using visual inputs versus using force inputs.

Table 5
Testing results of visual and force groups.

Groups	Time Step	Radius Error (m)	Success Rate (%)
Visual	2799	0.0972	77.9 %
Force	953	0.0321	92.5 %

sequence order [121]. This process often demands real-time adjustments due to changing conditions on the construction site, such as material variability and evolving project requirements [122]. In the context of HRI, effective sequence planning is essential not just for task execution but also for ensuring smooth collaboration between robots and human workers. The ability of robots to adapt their task sequence in response to human decisions and onsite changes greatly reduces time-consuming and error-prone manual planning, typically performed by construction workers [123].

At present, sequence planning is largely addressed by the assembly sequence planning [124] literature, which aims to identify the most efficient and cost-effective sequence of operations to assemble a product while considering various factors, such as resources, constraints, and goals [125]. Classical approaches such as the AND/OR graph [126] and the liaison graph [127] represent assembly operations as directed graphs with nodes representing parts and edges representing assembly operations. Matrix-based methods, such as the design structure matrix (DSM) [128] and the assembly incidence matrix (AIM) [129], use matrices to represent relationships between parts and assembly operations. Additionally, mathematical methods like integer programming (IP) [130], mixed-integer linear programming (MILP) [131], and constraint programming (CP) [132] have been used to model and solve ASP problems by formulating them as mathematical models with variables representing assembly operations and constraints representing precedence relationships and resource limitations. Heuristic methods, such as greedy algorithms [133], local search methods like simulated annealing [134], tabu search [135], and variable neighborhood search [136] have been employed to find good solutions efficiently by using heuristic rules to navigate the complex solution space of ASP problems. In recent years, approaches leveraging machine learning have emerged, such as genetic algorithms [137], ant colony optimization [138], particle swarm optimization [139], and artificial neural networks [140]. These methods are used to explore a wide search space and generate optimal or near-optimal solutions.

However, the traditional sequence planning paradigms often neglect the intricacies of real-world physical dynamics, leading to theoretically optimal sequences that may be unfeasible in a tangible environment. Embodied AI, underpinned by accurate physics engines, provides a foundational shift in this perspective. The physics engine facilitates a high-fidelity simulation environment, granting the ability to model, test, and validate sequence decisions under rigorous physical constraints. Within this environment, dynamic physics interactions can be accurately replicated, enabling a meticulous analysis of potential force interactions, torque requirements, and spatial constraints that a robot might encounter during real task execution. These simulations enable the identification and mitigation of potential pitfalls in a sequence. Furthermore, by incorporating physics simulations into the planning phase, we can transition from merely heuristic-based planning to a more holistic, physics-informed decision-making process. This methodology ensures that sequence decisions are not only algorithmically optimal but are also validated under a spectrum of real-world physical scenarios. In essence, embodied AI for sequence planning transcends the limitations of conventional algorithms by grounding decisions in tangible physics. This approach not only guarantees sequences that are computationally efficient but also those that stand the test of real-world dynamism and constraints, marking a significant stride in the evolution of robust robotic operations.

In the following test case, we show the implementation of using embodied simulation to solve sequential decision-making problems and control the robot arm to manipulate in a simulated platform with the help of Large Language Models (LLMs) [141,142]. This test case is set up as the pipeline installation which is a general task in the construction site. The robot is required to use pipes of the same size to create a pipeline from the given starting and ending points. We added three obstacles midway randomly, simulating unexpected constraints from real life, such as existing machines or faulty wall settings. The installation sequence was generated by fine-tuned LLMs [143–145]. With no unexpected obstacles, the LLMs could generate correct action sequences. However, when unexpected obstacles occurred, the LLMs could not determine a correct solution. We then provided the LLMs with an embodied environment that had contextual information such as the occurrence of obstacles. Specifically, we developed a token-based representation system that incorporated both symbolic and spatial information to describe the arrangement and state of the embodied environment. Each token consisted of a specific component, followed by its spatial coordinates, denoted as (x, y, z). For instance, pipe sections were represented as "PIPE (x,y,z)", the starting and ending points of the pipeline as "START (x,y,z)" and "END (x,y,z)", respectively. Unforeseen barriers, on the other hand, were denoted as "OBSTACLE_MACHINE (x, y,z)" or "OBSTACLE_WALLFAULT (x,y,z)" based on their type. Additionally, spatial relations were encoded using directional tokens like "LEFT_OF", "RIGHT_OF", "ABOVE", and "BELOW". To offer a clearer context, if a faulty wall setting was located at coordinates (3, 2, 1) and to the left of the second pipe situated at (4, 2, 1), it would be encoded as "OBSTACLE_WALLFAULT (3,2,1)"; LEFT_OF PIPE_2 (4,2,1)". Table 6 lists representative information used in our token system.

We presented these token sequences to the LLMs, giving them structured input that melded both symbolic and spatial data as shown in Fig. 10. With this enriched context, the LLMs planner was primed to process the scene, reason about the potential issues arising from the obstacles, and suggest alternative action sequences. These sequences guided the robot arm in maneuvering around the obstructions, ensuring the seamless installation of the pipeline.

3.7. Correction, rework or discard Decision-Making

The ability to make independent and spontaneous decisions towards mistakes in an assembly process is also essential for dexterity-capable robots in construction sites. We call it the correction, rework, or discard (CRD) decision-making problem. For construction (and many other industrial tasks), it is difficult to achieve a goal in a single attempt, particularly in complex and dynamic assembly tasks. When a mistake happens in the middle of an assembly operation, robots should be able to decide whether a corrective action or rework is more proper, or just simply discard the ongoing work. At present, most control methods command a direct halting when an unexpected scenario (such as mistakes) happens [146]. To perform tasks effectively and economically, robots need the capacity for self-correction, choosing between reworking, restarting, or discarding an action. This decision-making process, complex even for humans, involves considering technical, logistic, economic, and safety factors. For example, manufacturing literature has developed a level of repair analysis (LORA) framework that determines whether an item should be repaired, replaced, or discarded, guided by the considerations of cost and operational readiness requirements [147]. In the context of HRI, enhancing robots with CRD decision-making abilities ensures smoother, more autonomous collaboration, reducing reliance on human intervention for error.

After a comprehensive literature search, we could not find existing studies that presented explicit methods for making CRD decisions. Although the literature has explored machine learning approaches for equipping robots with the ability of corrective actions (e.g., [148,149]), these methods focus on how to better perform the corrective and rework actions after a CRD has been made. Instead, what we highlight here is the lack of a quantitative method or formulation for robots to make such a CRD decision without human intervention. As a result, we propose the

Table 6
Proposed token system for LLMs to understand physical conditions.

Token Type	Description	Example
Object Tokens: To inform LLMs the objects		
PIPE	Represents a pipe section.	**PIPE (5,2,1)**
START	Denotes the starting point of the pipeline.	**START (1,1,1)**
END	Denotes the ending point of the pipeline.	**END (10,1,1)**
OBSTACLE_MACHINE	Represents an unexpected machine-based obstacle.	**OBSTACLE_MACHINE (6,2,1)**
OBSTACLE_WALLFAULT	Represents an unexpected wall fault.	**OBSTACLE_WALLFAULT (7,2,1)**
Directional Tokens: Used to inform LLMs of current spatial conditions		
LEFT_OF	Indicates one object is to the left of another.	**OBSTACLE_MACHINE (6,2,1) LEFT_OF PIPE (7,2,1)**
RIGHT_OF	Indicates one object is to the right of another.	**PIPE(7,2,1) RIGHT_OF OBSTACLE_MACHINE(6,2,1)**
ABOVE	Indicates one object is above another.	**PIPE(5,3,1) ABOVE OBSTACLE_WALLFAULT (5,2,1)**
BELOW	Indicates one object is below another.	**PIPE(5,1,1) BELOW OBSTACLE_WALLFAULT (5,2,1)**
State Tokens: Used to inform LLMs the current system states		
INSTALLED	Indicates a pipe section has been installed.	**PIPE(5,2,1) INSTALLED**
UNINSTALLED	Indicates a pipe section hasn't been installed yet.	**PIPE(6,2,1) UNINSTALLED**
BLOCKED	Indicates a path or position is blocked by an obstacle.	**PATH(6,2,1) BLOCKED**
Action Tokens: Used for representing robot actions or decisions.		
MOVE_TO	Indicates the robot should move to a specific position.	**MOVE_TO(5,2,1)**
PICK_PIPE	Indicates the robot should pick up a specific pipe.	**PICK_PIPE(5,2,1)**
INSTALL_PIPE	Indicates the robot should install a pipe at a location.	**INSTALL_PIPE(6,2,1)**

use of embodied AI to provide a learning platform where robots can explore strategies for spontaneous CRD decision-making. By training AI within physically accurate simulators, these systems can learn from countless virtual experiments, discovering how different CRD strategies affect task outcomes under various conditions. Through reinforcement learning, the AI system can learn to select the most effective action – whether it is to correct, rework, or discard a task – in response to an error. This could potentially facilitate the development of construction robots that can respond adaptively and intelligently to unexpected scenarios, enhancing their efficiency and effectiveness in complex, dynamic assembly tasks. Imitation learning is also an effective approach for transferring human decision strategies to these systems, but the variability of human actions and strategies should be considered to generalize across different situations and workers. Given that this area of research is still primitive, no cases can be presented in this paper. Future research can explore methods for capturing and modeling the variability in human error correction and rework strategies, as well as developing algorithms that can learn from multiple demonstrations to achieve more robust and adaptable decision-making capabilities.

4. Discussion

The scalable adoption of construction robots has shown the potential to revolutionize the way motor-intensive construction assembly tasks are performed, leading to increased efficiency, safety, and productivity. One significant challenge relates to the development of dexterous robots, mainly manipulators, which can perform complex, high-precision tasks that are traditionally labor-intensive and prone to human error. This paper proposes a DEXBOT framework for designing dexterity-capable construction robots based on the principles of embodied AI, for better perception, planning, and control mechanisms to systematically improve robots' abilities to interact and collaborate with human workers effectively. Embodied AI combines AI with physics-based simulations, enabling robots to interact with human workers in a virtual yet physically accurate environment. This combination can enhance robots' ability to understand, learn from, and navigate complex real-world conditions, significantly improving their performance, adaptability, and decision-making capabilities which are crucial for sophisticated tasks in construction sites. The six fundamental steps within this framework include scene understanding, localization and motion planning, position-based control, force-based control, sequence planning, and the decision-making process concerning correction, rework, or discard. In each of these stages, embodied AI serves a pivotal role in transforming the current state-of-the-art practices.

First, embodied AI can significantly improve scene understanding

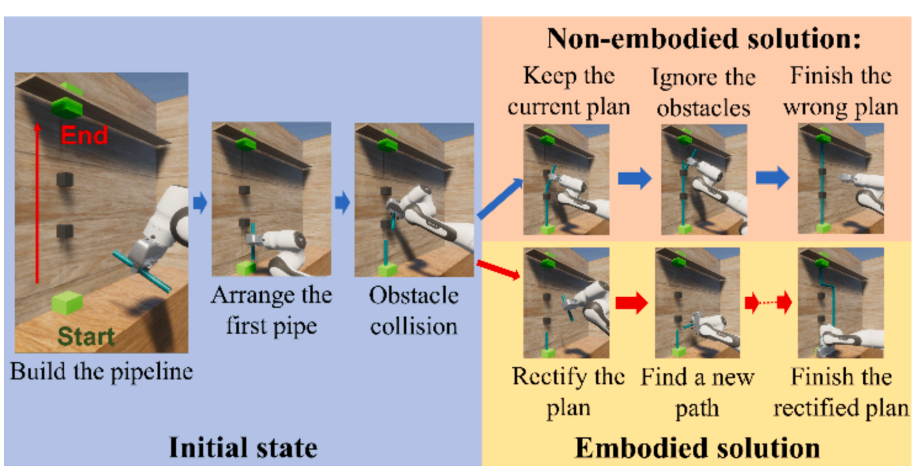

Fig. 10. The demonstration of the process by utilizing ChatGPT-4 to plan the pipeline and control the robot arm to perform the pipeline installation in the virtual environment.

(step 1) by integrating physics-based attributes into recognized objects within the raw reality capture data. For instance, this could involve leveraging machine learning techniques to classify and extract object properties such as mass, dimensions, and material characteristics from point clouds or other sensor data. The result is a richer, more comprehensive model of the environment that not only includes object identification but also physical properties, which can support more advanced interaction planning and decision-making capabilities in construction robots. Second, embodied AI has significant potential in enhancing robot localization and motion planning (step 2), as well as position-based control (step 3). It can provide a realistic simulation environment that closely mirrors real-world conditions. For example, movable objects can be identified to improve the proactivity of localization algorithms. Advanced motion planning algorithms can be trained and validated in this simulated environment, handling complex navigation tasks under different site conditions and obstacles. Similarly, embodied AI can help improve position-based controls, where the robot's end-effector follows a desired trajectory, by simulating different physical interactions and disturbances. Third, embodied AI can significantly contribute to the training and validation of robotic force-based controls (step 4). Physics simulation can be employed to model the forces involved in manipulating different objects, allowing the development, and testing of control strategies in a safe, controlled environment. Furthermore, creating environments that mimic real-world construction sites allows for the collection of meaningful human demonstration data, enabling the implementation of effective imitation learning techniques. Last but not least, embodied AI can also be utilized in the validation of sequence planning (step 5) and spontaneous decisions when any assembly mistake is observed (step 6). By providing a high-fidelity simulated environment, embodied AI can support machine learning models in identifying optimal assembly sequences and correcting courses when errors occur. With the high complexity of construction assembly tasks, the use of a simulated environment allows the sequence planner to be tested against a variety of scenarios, improving its generalization and robustness.

Note that the pipe installation task was specifically chosen for its embodiment of common construction challenges, including complex Spatial requirements, variability and adaptability, integration with existing systems and interaction with unstructured environments, making it an ideal candidate to showcase the DEXBOT framework's capabilities. Firstly, pipe installation often necessitates working within tightly constrained spaces and necessitates a high degree of spatial awareness. Robots must navigate these spaces while avoiding existing structures, which mirrors the spatial navigation challenges present in many other construction tasks. Secondly, the task involves handling materials with varying dimensions and specifications, requiring the robot to adapt its approach for different pipe sizes and materials. This variability demands a level of adaptability and decision-making that is crucial across construction tasks, where no two scenarios are identical. Thirdly, installing pipes involves integrating with existing systems (e.g., water, gas, HVAC), akin to how many construction tasks must consider and accommodate pre-existing structures and utilities. This aspect tests the robot's ability to work within a predefined framework, enhancing its applicability to diverse scenarios. In addition, unlike controlled environments, construction sites are dynamic and unpredictable. The pipe installation task, set within such an environment, challenges the robot to perform under variable conditions, including changes in lighting, weather, and the presence of unanticipated obstacles. Besides, the task also allows for the simulation of real-world time constraints and efficiency requirements, mirroring the pressures of actual construction projects where time is often a critical factor. By detailing the application of the DEXBOT framework to this specific scenario, we aim to highlight its versatility and effectiveness in addressing a wide spectrum of construction activities, underscoring the potential for scalable adoption of robotics in the construction sector.

This paper also provides test cases for the key steps of the proposed DEXBOT framework. Given these potentially transformative impacts, we encourage academic exploration of embodied AI's applications in construction, with more methodological and practical evidence. This allows for the deployment of robotic systems across various construction tasks and settings, making them more flexible assets for the industry.

Despite the promising advancements introduced by our novel construction robot technology, its deployment in real-world construction environments still has a thorough consideration of potential challenges and areas for future development. Construction sites present a unique set of conditions—ranging from harsh weather to highly variable and cluttered workspaces—that can significantly impact the operational effectiveness of robotic systems. The complexity of construction tasks, coupled with specialized processes inherent to the industry, requires robots to possess not only advanced dexterity and adaptability but also an intricate understanding of construction workflows and the ability to navigate them effectively. One of the principal challenges involves ensuring the robot's resilience to the diverse and often extreme conditions found on construction sites. Factors such as dust, moisture, and fluctuating temperatures can impede robotic sensors and machinery, necessitating the development of robust designs that safeguard against these environmental stresses. Moreover, the complexity and specificity of construction tasks demand that robots are equipped with sophisticated planning, decision-making, and execution capabilities. This includes the ability to adapt to unforeseen changes in the environment or task requirements, a critical feature for maintaining efficiency and safety on dynamic construction sites. Future research will focus on enhancing the robot's environmental robustness and its cognitive and mechanical adaptability. Additionally, integrating these robots into existing construction processes poses its own set of challenges. It is essential to develop seamless human-robot collaboration mechanisms, ensuring that robots can effectively collaborate with human counterparts without disrupting established workflows, thereby enhancing overall productivity and safety.

5. Conclusion

Construction tasks present a significant opportunity for robotic applications, yet there are still substantial challenges to overcome the limited dexterity capabilities of existing construction robotic methods. This paper proposes that the adoption of embodied AI will lead to transformative advancement for dexterity-capable construction robots to support sophisticated construction tasks by enhancing their level of intelligence and capabilities in multiple key areas of perception, planning, operations, and decision-making. The development of self-adapting, scalable robotic systems allow for the broad deployment of automation across a variety of construction scenarios.

While the development of embodied AI presents numerous opportunities for dexterity-capable construction robots, there are several challenges that must be addressed. First, as the embodied AI algorithms and control mechanisms for dexterity-capable construction robots grow in complexity, the computational demands increase significantly. This can lead to higher power consumption, slower processing times, and increased costs. To address this challenge, future research should focus on developing more efficient algorithms, leveraging edge computing, and exploring specialized hardware optimized for AI computations. Second, obtaining accurate and diverse training data is essential for developing effective embodied AI models. However, collecting and annotating large volumes of data can be both time-consuming and expensive. To address this challenge, researchers are encouraged to explore data augmentation and transfer learning techniques, which allow models to leverage pre-trained components, reducing the amount of data required for training. An industry-wide protocol for sharing data is a feasible solution. Additionally, synthetic data generation through simulations and procedural modeling can provide valuable training data in a more controlled and cost-effective manner. This is a natural benefit of embodied AI as existing literature in this area provides methods for generating physically accurate data that can be used to train robots.

Lastly, creating scalable and modular robotic systems that can adapt to various tasks and environments is crucial for widespread adoption in the construction industry. Possible solutions include soft, modular and reconfigurable robotic systems that provide adaptable and versatile robotic systems capable of interacting with diverse and unstructured environments.

The future agenda should also focus on the translation of these advancements from theoretical constructs to real-world applications is crucial. Field studies offer invaluable insights into how these strategies operate under real-world constraints and circumstances, highlighting practical challenges and limitations that may not be evident in controlled or simulated environments. We therefore advocate for robust field testing and data collection efforts, to allow us to iterate and improve upon these AI models in a data-driven and evidence-based manner. Moreover, bridging the gap between different academia and industry is critical to ensure that these advancements in embodied AI can be effectively adopted and deployed in real-world construction sites. This involves fostering partnerships between researchers, industry practitioners, and policy makers to align research directions with industry needs, address practical constraints and requirements, and facilitate the transition of these technologies into the marketplace. As dexterity-capable construction robots become more prevalent, ensuring the safety of human workers is the top priority. Developing reliable safety systems, such as real-time monitoring and collision avoidance, can help mitigate potential risks. Fostering trust between human workers and robots can be achieved through transparency in robot decision-making, providing humans with an understanding of the robot's intentions and actions, and allowing for more predictable and reliable interactions. By exploring new methods and practical standards for enabling dexterous robots for construction assembly tasks, we can pave the way for more capable automation for a safer, more efficient, and more productive construction industry.

CRediT authorship contribution statement

Hengxu You: Writing – original draft, Software, Formal analysis. **Tianyu Zhou:** Writing – review & editing, Methodology. **Qi Zhu:** Writing – review & editing, Methodology. **Yang Ye:** Writing – review & editing, Visualization, Validation. **Eric Jing Du:** Conceptualization, Project administration, Resources, Writing – review & editing.

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.

Data availability

Data will be made available on request.

Acknowledgements

This material is supported by the National Science Foundation (NSF) under grant 2024784. Any opinions, findings, conclusions, or recommendations expressed in this article are those of the authors and do not reflect the views of the NSF.

References

[1] T. Bock, Construction robotics, Auton. Robot. 22 (2007) 201–209.
[2] J.P. Sousa, P.A. Varela, P.F. Martins, Between manual and robotic approaches to brick construction in architecture, in: Real Time-Proceedings of the 33rd eCAADe Conference, 2015, pp. 361–370.
[3] T. Lei, Y. Rong, H. Wang, Y. Huang, M. Li, A review of vision-aided robotic welding, Comput. Ind. 123 (2020) 103326.
[4] S. Halder, K. Afsari, Robots in inspection and monitoring of buildings and infrastructure: a systematic review, Appl. Sci. 13 (2023) 2304.
[5] L. Ding, W. Jiang, Y. Zhou, C. Zhou, S. Liu, BIM-based task-level planning for robotic brick assembly through image-based 3D modeling, Adv. Eng. Inf. 43 (2020) 100993.
[6] L. Du, J. Yuan, S. Bao, R. Guan, W. Wan, Robotic replacement for disc cutters in tunnel boring machines, Autom. Constr. 140 (2022) 104369.
[7] J. Pereyras, Acceptability and technical feasibility of an electrical wiring installation trainer, (2020).
[8] A. Billard, D. Kragic, Trends and challenges in robot manipulation, Science 364 (2019) eaat8414.
[9] H. You, F. Xu, E. Du, Robot-based real-time point cloud digital twin modeling in augmented reality, Transforming Construction with Reality Capture Technologies (2022).
[10] M. Savva, A. Kadian, O. Maksymets, Y. Zhao, E. Wijmans, B. Jain, J. Straub, J. Liu, V. Koltun, J. Malik, Habitat: A platform for embodied ai research, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 9339–9347.
[11] J. Duan, S. Yu, H.L. Tan, H. Zhu, C. Tan, A survey of embodied ai: From simulators to research tasks, IEEE Transactions on Emerging Topics in Computational Intelligence 6 (2022) 230–244.
[12] R. Pfeifer, F. Iida, Embodied artificial intelligence: trends and challenges, Lect. Notes Comput. Sci (2004) 1–26.
[13] Q. Liu, Z. Liu, B. Xiong, W. Xu, Y. Liu, Deep reinforcement learning-based safe interaction for industrial human-robot collaboration using intrinsic reward function, Adv. Eng. Inf. 49 (2021) 101360.
[14] N. Melenbrink, J. Werfel, A. Menges, On-site autonomous construction robots: towards unsupervised building, Autom. Constr. 119 (2020) 103312.
[15] X. Xu, B.G. De Soto, in: On-Site Autonomous Construction Robots: A Review of Research Areas, Technologies, and Suggestions for Advancement, ISARC, IAARC Publications, 2020, pp. 385–392.
[16] A.-N. Sharkawy, Human-robot interaction: Applications, arXiv preprint arXiv: 2102.00928, (2021).
[17] L. Liu, A.J. Schoen, C. Henrichs, J. Li, B. Mutlu, Y. Zhang, R.G. Radwin, Human robot collaboration for enhancing work activities, Hum. Factors 66 (2024) 158–179.
[18] C.J. Conti, A.S. Varde, W. Wang, Task quality optimization in collaborative robotics, IEEE international Conference on Big Data (big data) IEEE (2020) 5652–5654.
[19] U. Othman, E. Yang, Human–robot collaborations in smart manufacturing environments: review and outlook, Sensors 23 (2023) 5663.
[20] G. Evangelou, N. Dimitropoulos, G. Michalos, S. Makris, An approach for task and action planning in human–robot collaborative cells using AI, Procedia Cirp 97 (2021) 476–481.
[21] A. Dai, C. Diller, M. Nießner, Sg-nn: Sparse generative neural networks for self-supervised scene completion of rgb-d scans, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 849–858.
[22] J. Hua, L. Zeng, G. Li, Z. Ju, Learning for a robot: deep reinforcement learning, imitation learning, transfer learning, Sensors 21 (2021) 1278.
[23] K. Arulkumaran, D.O. Lillrank, A pragmatic look at deep imitation learning, Asian Conference on Machine Learning, PMLR (2024) 58–73.
[24] J.M.D. Delgado, L. Oyedele, Robotics in construction: a critical review of the reinforcement learning and imitation learning paradigms, Adv. Eng. Inf. 54 (2022) 101787.
[25] C.-J. Liang, V.R. Kamat, C.C. Menassa, Teaching robots to perform quasi-repetitive construction tasks through human demonstration, Autom. Constr. 120 (2020) 103370.
[26] Y. Lu, J. Fu, G. Tucker, X. Pan, E. Bronstein, R. Roelofs, B. Sapp, B. White, A. Faust, S. Whiteson, Imitation is not enough: Robustifying imitation with reinforcement learning for challenging driving scenarios, in: 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2023, pp. 7553–7560.
[27] H. Li, W. Ma, H. Wang, G. Liu, X. Wen, Y. Zhang, M. Yang, G. Luo, G. Xie, C. Sun, A framework and method for human-robot cooperative safe control based on digital twin, Adv. Eng. Inf. 53 (2022) 101701.
[28] E. Magrini, F. Ferraguti, A.J. Ronga, F. Pini, A. De Luca, F. Leali, Human-robot coexistence and interaction in open industrial cells, Rob. Comput. Integr. Manuf. 61 (2020) 101846.
[29] R. Srivastava, S.H. Alsamhi, N. Murray, D. Devine, Shape memory alloy-based wearables: a review, and conceptual frameworks on HCI and HRI in Industry 4.0, Sensors 22 (2022) 6802.
[30] Y. Pan, L. Zhang, Roles of artificial intelligence in construction engineering and management: a critical review and future trends, Autom. Constr. 122 (2021) 103517.
[31] K.S. Saidi, T. Bock, C. Georgoulas, Robotics in construction, Springer, Springer handbook of robotics, 2016, pp. 1493–1520.
[32] J.P. Queralta, J. Taipalmaa, B.C. Pullinen, V.K. Sarker, T.N. Gia, H. Tenhunen, M. Gabbouj, J. Raitoharju, T. Westerlund, Collaborative multi-robot search and rescue: planning, coordination, perception, and active vision, IEEE Access 8 (2020) 191617–191643.
[33] C. Liu, J. Wu, X. Jiang, Y. Gu, L. Xie, Z. Huang, Automatic assembly of prefabricated components based on vision-guided robot, Autom. Constr. 162 (2024) 105385.
[34] X. Li, X. Jiang, Development of a robot system for applying putty on plastered walls, in: 2018 IEEE International Conference on Mechatronics and Automation (ICMA), 2018, pp. 1417–1422.

[35] E. Menendez, J.G. Victores, R. Montero, S. Martínez, C. Balaguer, Tunnel structural inspection and assessment using an autonomous robotic system, Autom. Constr. 87 (2018) 117–126.
[36] S. Siebert, J. Teizer, Mobile 3D mapping for surveying earthwork projects using an Unmanned Aerial Vehicle (UAV) system, Autom. Constr. 41 (2014) 1–14.
[37] V. Azamfirei, F. Psarommatis, Y. Lagrosen, Application of automation for in-line quality inspection, a zero-defect manufacturing approach, J. Manuf. Syst. 67 (2023) 1–22.
[38] X. Chen, H. Huang, Y. Liu, J. Li, M. Liu, Robot for automatic waste sorting on construction sites, Autom. Constr. 141 (2022) 104387.
[39] J. Park, P. Kim, Y.K. Cho, J. Kang, Framework for automated registration of UAV and UGV point clouds using local features in images, Autom. Constr. 98 (2019) 175–182.
[40] K. Asadi, A.K. Suresh, A. Ender, S. Gotad, S. Maniyar, S. Anand, M. Noghabaei, K. Han, E. Lobaton, T. Wu, An integrated UGV-UAV system for construction site data collection, Autom. Constr. 112 (2020) 103068.
[41] D. Hu, V.J. Gan, T. Wang, L. Ma, Multi-agent robotic system (MARS) for UAV-UGV path planning and automatic sensory data collection in cluttered environments, Build. Environ. 221 (2022) 109349.
[42] I. Jeong, Y. Jang, J. Park, Y.K. Cho, Motion planning of mobile robots for autonomous navigation on uneven ground surfaces, J. Comput. Civ. Eng. 35 (2021) 04021001.
[43] H.J. Wagner, M. Alvarez, O. Kyjanek, Z. Bhiri, M. Buck, A. Menges, Flexible and transportable robotic timber construction platform–TIM, Autom. Constr. 120 (2020) 103400.
[44] H. Bavle, J.L. Sanchez-Lopez, M. Shaheer, J. Civera, H. Voos, Situational graphs for robot navigation in structured indoor environments, IEEE Rob. Autom. Lett. 7 (2022) 9107–9114.
[45] K.M. Lundeen, V.R. Kamat, C.C. Menassa, W. McGee, Scene understanding for adaptive manipulation in robotized construction work, Autom. Constr. 82 (2017) 16–30.
[46] O. Kyjanek, B. Al Bahar, L. Vasey, B. Wannemacher, A. Menges, in: Implementation of an Augmented Reality AR Workflow for Human Robot Collaboration in Timber Prefabrication, International Association for Automation and Robotics in Construction (IAARC …, 2019, pp. 1223–1230.
[47] L.J. Tun, R.J. Rajapakse, K. Miyata, HaptWarp: implementation of bend, twist, and press manipulations for intuitive interaction in between the virtual and physical world, J. Robotics Netw. Artif. Life 8 (2022) 263–268.
[48] I. Dekker, K. Kellens, E. Demeester, Design and evaluation of an intuitive haptic teleoperation control system for 6-DoF industrial manipulators, Robotics 12 (2023) 54.
[49] C. Tiseo, W. Merkt, K.K. Babarahmati, W. Wolfslag, I. Havoutis, S. Vijayakumar, M. Mistry, Hapfic: an adaptive force/position controller for safe environment interaction in articulated systems, IEEE Trans. Neural Syst. Rehabil. Eng. 29 (2021) 1432–1440.
[50] V. De Simone, V. Di Pasquale, V. Giubileo, S. Miranda, Human-robot collaboration: an analysis of worker's performance, Procedia Comput. Sci. 200 (2022) 1540–1549.
[51] S. Baltrusch, F. Krause, A. de Vries, W. van Dijk, M. de Looze, What about the Human in Human Robot Collaboration? A Literature Review on HRC's effects on aspects of job quality, Ergonomics 65 (2022) 719–740.
[52] A. Kunic, R. Naboni, A. Kramberger, C. Schlette, Design and assembly automation of the robotic reversible timber beam, Autom. Constr. 123 (2021).
[53] M. Pan, T. Linner, W. Pan, H. Cheng, T. Bock, A framework of indicators for assessing construction automation and robotics in the sustainability context, J. Clean. Prod. 182 (2018).
[54] X. Ma, C. Mao, G. Liu, Can robots replace human beings?—assessment on the developmental potential of construction robot, Journal of Building Engineering 56 (2022) 104727.
[55] L. Bu, Y. Zhang, H. Liu, X. Yuan, J. Guo, S. Han, An IIoT-driven and AI-enabled framework for smart manufacturing system based on three-terminal collaborative platform, Adv. Eng. Inf. 50 (2021).
[56] C. Brosque, M. Fischer, A robot evaluation framework comparing on-site robots with traditional construction methods, Construction Robotics 6 (2022).
[57] J. Xu, T. Chen, L. Zlokapa, M. Foshey, W. Matusik, S. Sueda, P. Agrawal, An End-to-End Differentiable Framework for Contact-Aware Robot Design, 2021.
[58] D. Lee, S. Lee, N. Masoud, M.S. Krishnan, V.C. Li, Digital twin-driven deep reinforcement learning for adaptive task allocation in robotic construction, Adv. Eng. Inf. 53 (2022).
[59] B. Xiao, C. Chen, X. Yin, Recent advancements of robotics in construction, Autom. Constr. 144 (2022).
[60] X. Wang, S. Wang, C.C. Menassa, V.R. Kamat, W. McGee, Automatic high-level motion sequencing methods for enabling multi-tasking construction robots, Autom. Constr. 155 (2023).
[61] R. Naboni, A. Kunic, A computational framework for the design and robotic manufacturing of complex wood structures.
[62] S. Chang, M.-F.-F. Siu, H. Li, X. Luo, Evolution pathways of robotic technologies and applications in construction, Adv. Eng. Inf. 51 (2022).
[63] L. Foglia, R.A. Wilson, Embodied cognition, wiley interdisciplinary reviews, Cognit. Sci. 4 (2013) 319–325.
[64] R. Pfeifer, J. Bongard, How the body shapes the way we think: a new view of intelligence, MIT press2006.
[65] N. Roy, I. Posner, T. Barfoot, P. Beaudoin, Y. Bengio, J. Bohg, O. Brock, I. Depatie, D. Fox, D. Koditschek, From machine learning to robotics: challenges and opportunities for embodied intelligence, arXiv preprint arXiv:2110.15245 (2021).
[66] G. Cimolino, T.N. Graham, Two heads are better than one: A dimension space for unifying human and artificial intelligence in shared control, in: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, 2022, pp. 1–21.
[67] H. Su, W. Qi, J. Chen, C. Yang, J. Sandoval, M.A. Laribi, Recent advancements in multimodal human–robot interaction, Front. Neurorob. 17 (2023).
[68] D. Mukherjee, K. Gupta, L.H. Chang, H. Najjaran, A survey of robot learning strategies for human-robot collaboration in industrial settings, Rob. Comput. Integr. Manuf. 73 (2022) 102231.
[69] S. Li, P. Zheng, S. Liu, Z. Wang, X.V. Wang, L. Zheng, L. Wang, Proactive human–robot collaboration: mutual-cognitive, predictable, and self-organising perspectives, Rob. Comput. Integr. Manuf. 81 (2023) 102510.
[70] A. Fiske, P. Henningsen, A. Buyx, Your robot therapist will see you now: ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy, J. Med. Internet Res. 21 (2019) e13216.
[71] M.F. Stoelen, R. de Azambuja, B. López Rodríguez, F. Bonsignorio, A. Cangelosi, The GummiArm project: a replicable and variable-stiffness robot arm for experiments on embodied AI, Front. Neurorob. 16 (2022) 836772.
[72] B.R. Duffy, G. Joue, Intelligent robots: The question of embodiment, Proc. of the Brain-Machine Workshop, Citeseer, 2000.
[73] K. Asadi, H. Ramshankar, H. Pullagurla, A. Bhandare, S. Shanbhag, P. Mehta, S. Kundu, K. Han, E. Lobaton, T. Wu, Vision-based integrated mobile robotic system for real-time applications in construction, Autom. Constr. 96 (2018) 470–482.
[74] K. Kim, M. Peavy, BIM-based semantic building world modeling for robot task planning and execution in built environments, Autom. Constr. 138 (2022) 104247.
[75] F. Graf, J. Lindermayr, Ç. Odabaşi, M.F. Huber, Toward holistic scene understanding: a transfer of human scene perception to mobile robots, IEEE Rob. Autom. Mag. 29 (2022) 36–49.
[76] V. Lippiello, B. Siciliano, L. Villani, A position-based visual impedance control for robot manipulators, in: Proceedings 2007 IEEE International Conference on Robotics and Automation, 2007, pp. 2068–2073.
[77] F.J. Abu-Dakka, L. Rozo, D.G. Caldwell, Force-based variable impedance learning for robotic manipulation, Rob. Auton. Syst. 109 (2018) 156–167.
[78] D. Batra, A.X. Chang, S. Chernova, A.J. Davison, J. Deng, V. Koltun, S. Levine, J. Malik, I. Mordatch, R. Mottaghi, Rearrangement: A challenge for embodied ai, arXiv preprint arXiv:2011.01975, (2020).
[79] K. Van Wyk, M. Culleton, J. Falco, K. Kelly, Comparative peg-in-hole testing of a force-based manipulation controlled robotic hand, IEEE Trans. Rob. 34 (2018) 542–549.
[80] R. Ali, D. Kang, G. Suh, Y.-J. Cha, Real-time multiple damage mapping using autonomous UAV and deep faster region-based neural networks for GPS-denied structures, Autom. Constr. 130 (2021).
[81] J.A. Hama Kareem, B.I. Mohammed, S.A. Abdulwahab, Optimal materials handling equipment and defective product reduction skills in enhance overall production efficiency, SAGE Open 12 (2022).
[82] C.-L. Fan, Defect risk assessment using a hybrid machine learning method, J. Constr. Eng. Manag. 146 (2020).
[83] T. Taniguchi, S. Murata, M. Suzuki, D. Ognibene, P. Lanillos, E. Ugur, L. Jamone, T. Nakamura, A. Ciria, B. Lara, World models and predictive coding for cognitive and developmental robotics: frontiers and challenges, Adv. Rob. 37 (2023).
[84] C. Gan, S. Zhou, J. Schwartz, S. Alter, A. Bhandwaldar, D. Gutfreund, D.L.K. Yamins, J.J. DiCarlo, J. McDermott, A. Torralba, The threedworld transport challenge: A visually guided task-and-motion planning benchmark for physically realistic embodied ai, arXiv preprint arXiv:2103.14025, (2021).
[85] J. Jiang, Z. Huang, Z. Bi, X. Ma, G. Yu, State-of-the-Art control strategies for robotic PiH assembly, Rob. Comput. Integr. Manuf. 65 (2020).
[86] Q. Zhu, J. Du, Y. Shi, P. Wei, Neurobehavioral assessment of force feedback simulation in industrial robotic teleoperation, Autom. Constr. 126 (2021).
[87] H. Hoffmann, A. Jantsch, N.D. Dutt, Embodied self-aware computing systems, Proc. IEEE 108 (2020).
[88] L. Aymerich-Franch, E. Fosch-Villaronga, A self-guiding tool to conduct research with embodiment technologies responsibly, Frontiers in Robotics and AI 7 (2020).
[89] S.D. Das, V. Bain, P. Rakshit, Energy optimized robot arm path planning using differential evolution in dynamic environment, in: 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), 2018, pp. 1267–1272.
[90] J. Li, X. Zhang, J. Li, Y. Liu, J. Wang, Building and optimization of 3D semantic map based on Lidar and camera fusion, Neurocomputing 409 (2020) 394–407.
[91] A.S. Rao, M. Radanovic, Y. Liu, S. Hu, Y. Fang, K. Khoshelham, M. Palaniswami, T. Ngo, Real-time monitoring of construction sites: sensors, methods, and applications, Autom. Constr. 136 (2022) 104099.
[92] J. Crespo, J.C. Castillo, O.M. Mozos, R. Barber, Semantic information for robot navigation: a survey, Appl. Sci. 10 (2020) 497.
[93] L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, M. Pietikäinen, Deep learning for generic object detection: a survey, Int. J. Comput. Vis. 128 (2020) 261–318.
[94] W. Guo, G. Dong, C. Chen, M. Li, in: Learning Pushing Skills Using Object Detection and Deep Reinforcement Learning, 2019, pp. 469–474.
[95] T. Zhou, Q. Zhu, J. Du, Intuitive robot teleoperation for civil engineering operations with virtual reality and deep learning scene reconstruction, Adv. Eng. Inf. 46 (2020) 101170.
[96] D. Katainen, NVIDIA ISAAC AND ROBOT OPERATING SYSTEM 2 (2023).
[97] C.R. Qi, L. Yi, H. Su, L.J. Guibas, Pointnet++: deep hierarchical feature learning on point sets in a metric space, Adv. Neural Inf. Proces. Syst. 30 (2017).

[98] J. Bender, K. Erleben, J. Trinkle, Interactive simulation of rigid body dynamics in computer graphics, Wiley Online Library, Computer Graphics Forum, 2014, pp. 246–270.
[99] J. Zhang, X. Yang, W. Wang, J. Guan, L. Ding, V.C. Lee, Automated guided vehicles and autonomous mobile robots for recognition and tracking in civil engineering, Autom. Constr. 146 (2023) 104699.
[100] X. Xiao, B. Liu, G. Warnell, P. Stone, Motion planning and control for mobile robot navigation using machine learning: a survey, Auton. Robot. 46 (2022) 569–597.
[101] T.T. Mac, C. Copot, D.T. Tran, R. De Keyser, Heuristic approaches in robot path planning: a survey, Rob. Auton. Syst. 86 (2016) 13–28.
[102] A. Nessa, B. Adhikari, F. Hussain, X.N. Fernando, A survey of machine learning for indoor positioning, IEEE Access 8 (2020) 214945–214965.
[103] H. Taheri, Z.C. Xia, SLAM; definition and evolution, Eng. Appl. Artif. Intel. 97 (2021) 104032.
[104] S. Srivastava, C. Li, M. Lingelbach, R. Martín-Martín, F. Xia, K.E. Vainio, Z. Lian, C. Gokmen, S. Buch, K. Liu, Behavior: Benchmark for everyday household activities in virtual, interactive, and ecological environments, Conference on Robot Learning, PMLR, 2022, pp. 477-490.
[105] L. Pérez, E. Diez, R. Usamentiaga, D.F. García, Industrial robot control and operator training using virtual reality interfaces, Comput. Ind. 109 (2019) 114–120.
[106] A. Juliani, V.-P. Berges, E. Teng, A. Cohen, J. Harper, C. Elion, C. Goy, Y. Gao, H. Henry, M. Mattar, Unity: A general platform for intelligent agents, arXiv preprint arXiv:1809.02627, (2018).
[107] H. Shi, R. Li, X. Bai, Y. Zhang, L. Min, D. Wang, X. Lu, Y. Yan, Y. Lei, A review for control theory and condition monitoring on construction robots, J. Field Rob. 40 (2023) 934–954.
[108] M. Matulis, C. Harvey, A robot arm digital twin utilising reinforcement learning, Comput. Graph. 95 (2021) 106–114.
[109] H. Yin, A. Varava, D. Kragic, Modeling, learning, perception, and control methods for deformable object manipulation, Science, Robotics 6 (2021) eabd8803.
[110] T. Zhou, P. Xia, Y. Ye, J. Du, Embodied robot teleoperation based on high-fidelity visual-haptic simulator: pipe-fitting example, J. Constr. Eng. Manag. 149 (2023) 04023129.
[111] T. Johannink, S. Bahl, A. Nair, J. Luo, A. Kumar, M. Loskyll, J.A. Ojea, E. Solowjow, S. Levine, Residual reinforcement learning for robot control, in: 2019 International Conference on Robotics and Automation (ICRA), IEEE, 2019, pp. 6023–6029.
[112] S. Proia, R. Carli, G. Cavone, M. Dotoli, A Literature Review on Control Techniques for Collaborative Robotics in Industrial Applications, in: 2021 IEEE 17th International Conference on Automation Science and Engineering (CASE), IEEE, 2021, pp. 591–596.
[113] P. Song, Y. Yu, X. Zhang, Impedance control of robots: an overview, 2nd international Conference on Cybernetics, Robotics and Control (CRC) IEEE 2017 (2017) 51–55.
[114] M.Y. Cao, S. Laws, F.R. y Baena, Six-axis force/torque sensors for robotics applications: a review, IEEE Sens. J. 21 (2021) 27238–27251.
[115] Y. Deng, G. Wang, X. Yue, K. Zhou, A review of robot grinding and polishing force control mode, in: 2022 IEEE International Conference on Mechatronics and Automation (ICMA), IEEE, 2022, pp. 1413–1418.
[116] G. Raiola, C.A. Cardenas, T.S. Tadele, T. De Vries, S. Stramigioli, Development of a safety-and energy-aware impedance controller for collaborative robots, IEEE Rob. Autom. Lett. 3 (2018) 1237–1244.
[117] Z. Wang, L. Zou, X. Su, G. Luo, R. Li, Y. Huang, Hybrid force/position control in workspace of robotic manipulator in uncertain environments based on adaptive fuzzy control, Rob. Auton. Syst. 145 (2021) 103870.
[118] K. Chin, T. Hellebrekers, C. Majidi, Machine learning for soft robotic sensing and control, Advanced Intelligent Systems 2 (2020) 1900171.
[119] Y. Long, W. Wei, T. Huang, Y. Wang, Q. Dou, Human-in-the-loop embodied intelligence with interactive simulation environment for surgical robot learning, IEEE Rob. Autom. Lett. (2023).
[120] A. Ghadirzadeh, A. Maki, D. Kragic, M. Björkman, Deep predictive policy training using reinforcement learning, in: 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2017, pp. 2351–2358.
[121] Y. Shi, J. Du, D.A. Worthy, The impact of engineering information formats on learning and execution of construction operations: A virtual reality pipe maintenance experiment, Autom. Constr. 119 (2020) 103367.
[122] Z. Abuwarda, T. Hegazy, Work-package planning and schedule optimization for projects with evolving constraints, J. Comput. Civ. Eng. 30 (2016) 04016022.
[123] T. Kim, Y.-W. Kim, H. Cho, Dynamic production scheduling model under due date uncertainty in precast concrete construction, J. Clean. Prod. 257 (2020) 120527.
[124] I.A. Kazerouni, L. Fitzgerald, G. Dooly, D. Toal, A survey of state-of-the-art on visual SLAM, Expert Syst. Appl. 205 (2022) 117734.
[125] S. Alatartsev, S. Stellmacher, F. Ortmeier, Robotic task sequencing problem: a survey, J. Intell. Rob. Syst. 80 (2015) 279–298.
[126] T. Cao, A.C. Sanderson, AND/OR net representation for robotic task sequence planning, IEEE Trans. Syst. Man Cybern. Part C (applications and Reviews) 28 (1998) 204–218.
[127] Y. Xing, G. Chen, X. Lai, S. Jin, J. Zhou, Assembly sequence planning of automobile body components based on liaison graph, Assem. Autom. 27 (2007) 157–164.
[128] A. Yassine, An introduction to modeling and analyzing complex product development processes using the design structure matrix (DSM) method, Urbana 51 (2004) 1–17.
[129] I.-M. Chen, G. Yang, Configuration independent kinematics for modular robots, Proceedings of IEEE International Conference on Robotics and Automation, IEEE (1996) 1440–1445.
[130] T. Vossen, M. Ball, A. Lotem, D. Nau, Applying integer programming to AI planning, Knowl. Eng. Rev. 15 (2000) 85–100.
[131] J.P. Garcia-Sabater, J. Maheut, J.J. Garcia-Sabater, A two-stage sequential planning scheme for integrated operations planning and scheduling system using MILP: the case of an engine assembler, Flex. Serv. Manuf. J. 24 (2012) 171–209.
[132] N. Sundström, O. Wigström, P. Falkman, B. Lennartson, Optimization of operation sequences using constraint programming, IFAC Proceedings Volumes 45 (2012) 1580–1585.
[133] Z.-F. Liu, D. Hu, X. gao, J.-D. zhang, Product disassembly sequence planning based on greedy algorithm, China Mechanical Engineering 22 (2011) 2162.
[134] H. Shan, S. Li, D. Gong, P. Lou, Genetic simulated annealing algorithm-based assembly sequence planning, 2006 International Technology and Innovation Conference (ITIC 2006) IET 2006 (2006) 1573–1579.
[135] W. Li, S. Ong, A. Nee, Optimization of process plans using a constraint-based tabu search approach, Int. J. Prod. Res. 42 (2004) 1955–1985.
[136] H. Xia, X. Li, L. Gao, A hybrid genetic algorithm with variable neighborhood search for dynamic integrated process planning and scheduling, Comput. Ind. Eng. 102 (2016) 99–112.
[137] H.-E. Tseng, C.-C. Chang, S.-C. Lee, Y.-M. Huang, A block-based genetic algorithm for disassembly sequence planning, Expert Syst. Appl. 96 (2018) 492–505.
[138] Z. Han, Y. Wang, D. Tian, Ant colony optimization for assembly sequence planning based on parameters optimization, Frontiers of, Mech. Eng. 16 (2021) 393–409.
[139] L.A. Bewoor, V.C. Prakash, S.U. Sapkal, Production scheduling optimization in foundry using hybrid particle swarm optimization algorithm, Procedia Manuf. 22 (2018) 57–64.
[140] M. Suszyński, K. Peta, Assembly sequence planning using artificial neural networks for mechanical parts based on selected criteria, Appl. Sci. 11 (2021) 10414.
[141] H. Touvron, T. Lavril, G. Izacard, X. Martinet, M.-A. Lachaux, T. Lacroix, B. Rozière, N. Goyal, E. Hambro, F. Azhar, (2023) Llama: Open and efficient foundation language models arXiv preprint arXiv:2302.13971.
[142] R. OpenAI, GPT-4 technical report, arXiv, (2023) 2303.08774.
[143] Y. Ye, H. You, J. Du, Improved trust in human-robot collaboration with ChatGPT, IEEE Access (2023).
[144] H. You, Y. Ye, T. Zhou, Q. Zhu, J. Du, Robot-Enabled Construction Assembly with Automated Sequence Planning based on ChatGPT: RoboGPT, arXiv preprint arXiv:2304.11018, (2023).
[145] F. Xu, T. Nguyen, J. Du, Augmented reality for maintenance tasks with ChatGPT for automated text-to-action, J. Constr. Eng. Manag. (2023).
[146] C. Lopes-Dias, A.I. Sburlea, G.R. Müller-Putz, Online asynchronous decoding of error-related potentials during the continuous control of a robot, Sci. Rep. 9 (2019) 17596.
[147] R.J. Basten, J.M.J. Schutten, M.C. Van Der Heijden, An efficient model formulation for level of repair analysis, Ann. Oper. Res. 172 (2009) 119–142.
[148] C. Celemin, G. Maeda, J. Ruiz-del-Solar, J. Peters, J. Kober, Reinforcement learning of motor skills using policy search and human corrective advice, The International Journal of Robotics Research 38 (2019) 1560–1580.
[149] C. Celemin, J. Ruiz-del-Solar, J. Kober, A fast hybrid reinforcement learning framework with human corrective feedback, Auton. Robot. 43 (2019) 1173–1186.