Cloud-Based Hierarchical Imitation Learning for Scalable Transfer of Construction Skills from Human Workers to Assisting Robots

Hongrui Yu, S.M.ASCE[1]; Vineet R. Kamat, Ph.D., F.ASCE[2]; and Carol C. Menassa, Ph.D., F.ASCE[3]

Abstract: Assigning repetitive and physically demanding construction tasks to robots can alleviate human workers' exposure to occupational injuries, which often result in significant downtime or premature retirement. However, the successful delegation of construction tasks and the achievement of high-quality robot-constructed work requires transferring necessary dexterous and adaptive construction craft skills from workers to robots. Predefined motion planning scripts tend to generate rigid and collision-prone robotic behaviors in unstructured construction site environments. In contrast, imitation learning (IL) offers a more robust and flexible skill transfer scheme. However, the majority of IL algorithms rely on human workers repeatedly demonstrating task performance at full scale, which can be counterproductive and infeasible in the case of construction work. To address this concern, in this paper, we propose an immersive and Cloud Robotics-based virtual demonstration framework that serves two primary purposes. First, it digitalizes the demonstration process, eliminating the need for repetitive physical manipulation of heavy construction objects. Second, it employs a federated collection of reusable demonstrations that are transferable for similar tasks in the future and can, consequently, reduce the requirement for repetitive illustration of tasks by human agents. In addition, to enhance the trustworthiness, explainability, and ethical soundness of the robot training, this framework utilizes a hierarchical imitation learning (HIL) model to decompose human manipulation skills into sequential and reactive subskills. These two layers of skills are represented by deep generative models; these models enable adaptive control of robot action. The proposed framework has the potential to mitigate technical adoption barriers and facilitate the practical deployment of full-scale construction robots to perform a variety of tasks with human supervision. By delegating the physical strains of construction work to human-trained robots, this framework promotes the inclusion of workers with diverse physical capabilities and educational backgrounds within the construction industry. **DOI: 10.1061/JCCEE5.CPENG-5731.**

Introduction

The construction industry currently faces substantial challenges in its workforce (Delvinne et al. 2020). An estimated labor shortage of 430,000 construction workers (BLS 2023) compounds the industry's difficulties, which are further amplified by a forecasted increasing demand for labor (Wilder 2013; AGC 2016; Delvinne et al. 2020). For labor-intensive construction work, this shortage results directly in project delays and increased costs (Sokas et al. 2019). According to RSMeans, burgeoning labor shortages contributed to a 10% increase in total construction costs in 2016 (RSMeans 2016).

Construction robots can play a crucial role in addressing the workforce gap and mitigating the challenges posed by labor shortages (Lundeen et al. 2018; Wang et al. 2021a; Brosque and Fischer 2022). Robots possess superior physical capabilities and, therefore, excel in handling heavy and repetitive construction tasks (Liang et al. 2021, 2023). Furthermore, they are less prone to physical fatigue and cognitive lapses (Shayesteh and Jebelli 2021; Lee et al. 2022). By deploying construction robots, significant improvements can be achieved in construction productivity, leading to reduced delays and construction costs (Ryu et al. 2021; Pan and Pan 2020).

Human-Robot Craft Skill Learning Comparison

However, unstructured construction site environments pose inherent challenges for robots performing construction tasks (Feng et al. 2016; Liang and Cheng 2023). Accomplishing any complex motion planning for construction robots requires intricate motion control scripts (Cai et al. 2023; Zhu et al. 2022). Modeling and learning adaptive construction skills is increasingly challenging due to the abundance of information that influences the reasoning behind specific actions (Makondo et al. 2015; Xu et al. 2020). For example, in Lundeen et al. (2017), a workpiece's geometric configuration was reconstructed by carefully calibrating sensors to establish the correlation of choice of robot actions to the environmental observations. Ignoring such correlations and the accompanying spatial context tends to limit a robot to generating inflexible and inadequate motions (Haddadin et al. 2017). In a dynamic and unstructured environment like a construction site, relying solely on rigid robot control schemes significantly increases the risk of a robot getting stuck (i.e., stalling) or colliding with adjacent objects or human workers (i.e., interfering) (Zhu et al. 2022; Sun et al. 2023).

In contrast, human apprentices (junior construction workers) acquire dexterous manipulation skills and complex construction techniques organically by observing experienced workers (Sings et al. 2017; Liang et al. 2021). Through their observations, apprentices

[1]Ph.D. Candidate, Dept. of Civil and Environmental Engineering, Univ. of Michigan, Ann Arbor, MI 48109. Email: yhongrui@umich.edu

[2]Professor, Dept. of Civil and Environmental Engineering, Univ. of Michigan, Ann Arbor, MI 48109 (corresponding author). ORCID: https://orcid.org/0000-0003-0788-5588. Email: vkamat@umich.edu

[3]Professor, Dept. of Civil and Environmental Engineering, Univ. of Michigan, Ann Arbor, MI 48109. Email: menassa@umich.edu

Note. This manuscript was submitted on September 20, 2023; approved on January 3, 2024; published online on April 17, 2024. Discussion period open until September 17, 2024; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Computing in Civil Engineering***, © ASCE, ISSN 0887-3801.

J. Comput. Civ. Eng., 2024, 38(4): 04024019

instinctively establish correlations between sequential actions and the underlying motivations that drive them. These motivations are influenced by contextual information derived from the environment, previous actions, and overall work progress (Yu et al. 2023b; Liang et al. 2021; Lee et al. 2022; Huang et al. 2023). Consequently, workers naturally transform their observations into a cohesive set of actions guided by environmental cues and the sequence of preceding actions.

Skill Transfer from Humans to Robots with Imitation Learning

Replicating the learning process of humans, robot imitation learning (IL) (i.e., learning from demonstration) models offer a flexible and effective approach to motion control. These models employ neural networks or sequential chains to map complex environmental observations to desired actions (Liang et al. 2021; Huang et al. 2023; Wang et al. 2023a). For instance, a deep neural network can represent expected robot action policies based on various factors, including observed equipment status, worker states, task progress, and workpiece state. Reinforcement learning (RL) is also often employed to simulate the sequential decision-making processes of human workers, forming the deep reinforcement learning (DRL) framework, which is a prevalent approach in robot learning (DelPreto et al. 2020).

Low-Workload and Ethical Imitation Learning

However, applying IL to construction robots presents additional challenges for two primary reasons.

First, construction materials tend to be heavy, and repetitive manipulation during demonstrations can impose increased physical strain on construction workers. To address this issue, in this paper we propose a high-fidelity digital training environment for construction (DTEC) that enables human workers to naturally demonstrate installation tasks by performing construction tasks with digital twin models in virtual reality (VR). The DTEC connects with building information modeling (BIM) modules, which are capable of being automatically synchronized with site conditions (Fang and Cho 2016; Du et al. 2018; Wang et al. 2023b). This interface also helps increase worker trust in construction robots and their self-efficacy (Adami et al. 2022). In addition, in this study, we propose a federated construction skill cloud database that can leverage previous knowledge and demonstrations collected from diverse task environments in order to reduce the number of required future demonstrations. Such federated data collection allows for crowdsourced demonstrations from workspaces representative of various temporal and spatial contexts, enabling continuous learning and enhancing scalability in skill transfer between human workers and assisting robots (DelPreto et al. 2020).

Second, with regard to the ethical and responsible aspects of artificial intelligence (AI), the traditional RL-based structure lacks transparency and explainability (Gunning 2017). From a human perspective, the robot learning model appears as a black box and nontransparent model. This lack of transparency may increase the cognitive load on construction workers who lack programming expertise and reduce their trust in and willingness to collaborate with robots (Gunning 2017; Park et al. 2023). To address this issue, an explicable construction task decomposition framework is proposed in this paper. It involves decomposing construction skills into sequential and reactive skills. A hierarchical imitation learning (HIL) model is employed to ground the decomposed skills and translate them into robot control instructions. With improved explainability, the robot learning model will appear more trustworthy and easier for human workers without programming expertise to use (Hamon et al. 2020). In the long run, the model will result in a more sustainable human–robot relationships and enhance the equity and inclusion of workers with diverse educational background.

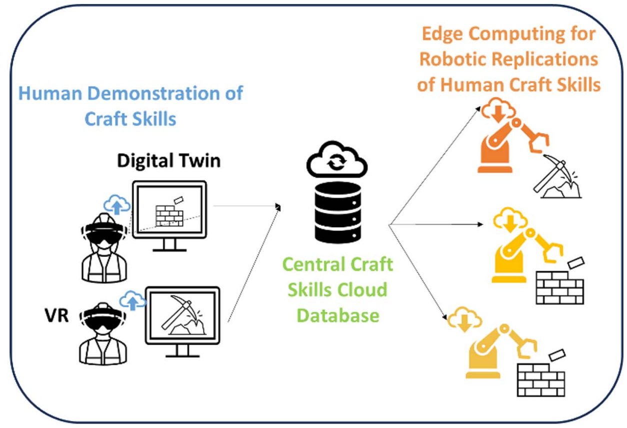

Fig. 1. Cloud-based IL framework for construction robots.

Fig. 1 summarizes this paper's technical contributions in enabling skills transfers from human workers to assisting construction robots, which include:

1. An immersive virtual demonstration environment (i.e., DTEC), comprising a digital twin, VR, and wireless data communications integrated with the robot operating system (ROS). This training environment enables automated capture of workers' motions and records them as demonstrations for robots, minimizing the physical workload on workers.
2. A cloud robotics framework designed to store, process, and reuse the demonstrations collected from the immersive virtual training environment across diverse task scenarios. This federated framework reduces the necessity for new demonstrations, thereby enhancing the scalability and sustainability of construction robot learning.
3. An explainable representation and transfer scheme for construction skills that effectively translates human manipulation motions into a hierarchical robot learning model capable of decoding both reactive and sequential skills. By incorporating this knowledge back into the cloud database for storage, the scalability and reusability of construction robot programming are further amplified.

The remainder of the paper is organized as follows. First, we provide an introductory literature review of the novel concepts of HIL and cloud robotics, demonstrating its suitability for application in the construction industry. Second, we present the methodology for the demonstration collection and skill decomposition system based on Cloud Robotics. Third, we present experiments conducted to evaluate the performance of the system, focusing on its ability to reduce the human effort required in both the demonstration and programming phases.

Literature Review

The transfer of skills between human workers and robots has been widely studied in order to enable robots to adapt to diverse construction environments (Lundeen et al. 2018; Yu et al. 2023b). While predefined motion scripts have been considered viable solutions for tasks such as bricklaying (Wos and Dindorf 2023), such an approach is unable to accommodate significant variations in the environment and the corresponding adjustments needed in robot actions (Shi et al. 2023). To overcome these limitations, there has been a growing focus on leveraging deep learning and probabilistic models to mimic the adaptive decision-making process of humans in response to environmental observations.

As these models evolve to become more accessible to human workers, there is an increasing demand for scalability, ease-of-use, and explainability (Gunning 2017; Park et al. 2023). The framework proposed in this paper addresses these requirements through the integration of cloud robotics and a hierarchical human skill representation scheme. This section provides an overview of both methods, highlighting their potential application in the construction field. At the end of each subsection, we present a summary of the current research and identify knowledge gaps pertaining to the effective implementation of these concepts on construction sites. This information establishes the motivation for the proposed framework, which is subsequently described in the methodology section.

Explainable Human–Robot Master-Apprentice Relationship

The seamless transfer of skills between construction workers and robots is essential for achieving high-quality corobotized construction tasks (Xu et al. 2020; Shi et al. 2023). Construction workers possess dexterity and specialized expertise, enabling them to avoid collisions and effectively address construction-related challenges (Liang et al. 2021). However, given that programming and robotics training are not typically part of their skill set, programming robots is an exceedingly difficult task (Liu et al. 2019). Liang et al. (2020) proposed construction robot IL (i.e., learning from demonstrations) as an approach to address this problem. This approach to human–robot collaboration requires human workers to demonstrate the natural execution process of tasks. Demonstrations are captured by video recording and then correlated with environmental observations using a DRL model. However, human workers must repetitively manipulate heavy construction materials to provide sufficient demonstration data. Huang et al. (2023) explored the use of digital demonstrations to alleviate the physical strain associated with manipulating heavy construction materials. Although this study creatively improved the demonstration environment, its kinesthetic demonstration technique, which involved the human moving the robot end-effector in a VR environment to showcase the desired trajectory, heavily depended on the robot's specific structure, limiting generalizability to different robot types. To extend application to a wider range of robot, in this study we designed demonstrations to represent the installation trajectories of tasks rather than the joint states of a specific robot.

Furthermore, as mentioned in the introduction, a common issue with robot learning models is their lack of transparency and explainability. This lack of clarity and understanding may lead to reduced trust and perceived validity among human workers (Gunning 2017). HIL addresses this concern by mimicking the natural way in which humans break down and simplify problems (Zhang et al. 2021). In HIL, high-level planning serves the purpose of task decomposition and connect the subtasks to a sequence of actions (Abdo et al. 2012). Low-level policies link task execution steps with environmental observations, such as task progress, object states, and robot states (Xie et al. 2020). For example, Zhang et al. (2021) decoded the robot pouring tasks and categorize them to three subtasks—phase, state, and action—to improve task performance, adaptability, and manipulability. Wang et al. (2021b) used HIL to teach robots to perform low-clearance insertion assembly tasks with increased sample efficiencies.

By employing two levels of policies, HIL achieves enhanced learning efficiency and reduces the need for a large number of demonstrations in long-horizon and complex tasks due to compounding errors (Kase et al. 2020; Yu et al. 2023a; Ross et al. 2011). For example, Hayashi et al. (2022) reduced training time by 20% by decomposing tasks into a planning layer and a motion primitive layer to ground the elemental motions. There are additional advantages. First, HIL reduces the errors and uncertainties that accumulate during the robot behavior cloning process, improving learning performance (Zhang et al. 2021). Second, it enables the use of conditional probabilities from previously observed actions to enhance confidence and simplify the inference of the current state from observations. This mitigates perceptual aliasing, in which similar environmental observations may lead to confusion between different task steps that serve distinct functions (Kase et al. 2020). For instance, when a robot aims to install three drywall panels consecutively and environmental observations indicate that the robot is at the location of the second panel, there are two possible action options, depending on task progress. Knowing the status of the first board installation makes the decision-making process considerably more straightforward.

Hierarchical task decomposition has been proposed and explored in the context of construction problems in prior studies. For instance, Wu et al. (2022) proposed a four-level hierarchy to decompose bricklaying tasks into two levels of subtasks: activities and actions. However, there is limited research available that connects such knowledge with robot learning models. It is still unclear how to decompose construction tasks and clearly represent them with deep robot learning models for robot control. To address this problem, in this study we explored how to decompose tasks with two layers: a layer of sequential actions linking the different subtasks and a layer of reactive skills that correlates environmental observations with choices of subtasks. The details of this technical approach are described in the methodology section.

Skill Transfer Scalability with Cloud Robotics

With HIL, humans can seamlessly transfer skills to robots. However, the current state of robot IL schemes is primarily need-based, requiring human operators to program robots from scratch whenever a new task arises. Such repetitive programming activities impose an additional workload on human workers. A common approach to address this challenge is to employ a knowledge database for the sustainable reuse of programmed instructions. For instance, Karp et al. (1994) proposed building a robot knowledge storage database and utilizing a database management system (DBMS) to improve communication efficiency. Bandera et al. (2010) introduced a robot knowledge database architecture that combines perceptive information, reflexive behaviors, and known actions.

Compared to local databases, cloud-based databases offer greater convenience for sharing and communication between multiple agents. As cloud-based systems gain popularity, more studies are being conducted to explore effective database architectural designs to adapt to different tasks. Perceptual information, human demonstrations, and robot executions are commonly stored. For dexterous tasks, Hsiao and Lozano-Perez (2006) recorded keyframes of successful grasp trajectories to teach robots precise grasp tasks with an accuracy of 92%. Zhang et al. (2021) illustrated how robots can be instructed to perform pouring tasks using demonstrations that include background information, task objects, and task planning policies.

In addition to human-demonstrated trajectories, supervisory commands can also serve as demonstrations and be stored in a robot dexterous manipulation learning database. Yamada et al. (2001) recorded natural language instructions from humans and the corresponding robot motion trajectories, manipulation objects, and elementary motions for a domestic robot operating under human supervision. Neto and Mendes (2013) employed a knowledge database to store analyzed parameters of movement commands, entire motion sequences, velocities, accelerations, blending behaviors, and signal-switching points.

Furthermore, for the hierarchical decomposition of more complex tasks, Kyrarini et al. (2019) proposed storing atomized tasks, associated object states, and learned Gaussian mixture models (GMMs) representing trajectories. Liu et al. (2022) collected both human-demonstrated trajectories and robot execution joint states, gripper dimension parameters, and gripper pose data to teach robots assembly and insertion tasks, effectively reducing the workload for humans. Wang et al. (2020) used stored dual-arm robot pose and point cloud models to enhance perception and improve model rebuilding accuracy for assembly tasks.

However, despite these initial studies, it is unclear what data or demonstrations are needed to compose a feasible cloud database capable of supporting local robot task execution decision making and planning for both heavy and dexterous construction tasks. Moreover, there is limited understanding of the impact of data stored in a cloud database on future task execution in unexpected environments. In this paper, we propose the dual storage of knowledge and raw demonstrations. This process was tested and demonstrated using a construction task case study. Previously stored data were added to the robot learning model for new tasks situations in order to investigate the usefulness of cloud storage in robot learning. The proposed approach is further described in the methodology section.

Cloud Robotics in Construction

In the civil sector, cloud databases and computing are widely used for natural disaster response, worker safety management, waste minimization, building management, and project management informatics (Jiao et al. 2013; Rawai et al. 2013; Balaji et al. 2016; Xu et al. 2018; Wan et al. 2020; Bello et al. 2021; Kohler et al. 2022; Deng et al. 2023). Nonetheless, their application in robotic construction has been relatively limited. Considering the effectiveness of cloud databases in exchanging information, building a cloud robotics scheme for construction robot training offers significant promise. In this paper, we propose a federated construction robot learning scheme to leverage previously stored demonstrations and information. Using edge computing and HIL, local robot control scripts can be retrieved on an as-needed basis to optimize the efficiency of computation resources. The proposed approach can reduce workers' physical demonstration workload and the mental workload when they collaborate with robots.

Research Methodology

This paper introduces a novel federated learning framework for construction robots that utilizes cloud robotics technology. The framework employs an immersive VR demonstration interface that is connected to a cloud database housing both crowdsourced raw demonstration data and explanatory knowledge on hierarchical construction skill decompositions. Within the cloud edge, a HIL algorithm is utilized to decode and model human adaptive skills from the demonstrations and data, which are subsequently replicated to guide the operations of construction robots. The following sections elaborate the technical details of each component. In addition, a case study of ceiling installation was used to illustrate how the learning algorithms can be applied to construction tasks.

Cloud-Robotics Learning Workflow Overview

Introducing new methods for robotic construction and robot programming inevitably brings changes to construction practice and workflow. The proposed workflow, adapting to the cloud-robotics scheme, is shown in Fig. 2.

In this scheme, a robot first recognizes task and target information by scanning AprilTag markers and BIM database to retrieve the task object's dimensions, material specifications, and locations (Aryan et al. 2021; Wang et al. 2023b). Subsequently, the robot accesses related demonstrations available in the cloud database. If sufficient data exist, the robot downloads them to the edge node and commences imitation learning to derive expected installation trajectories.

If cloud data are insufficient, the target information is visualized simultaneously by the DTEC to create a digital training setting for human workers to generate demonstrations. Relevant demonstrations from the cloud database, with heterogeneous tasks with similar

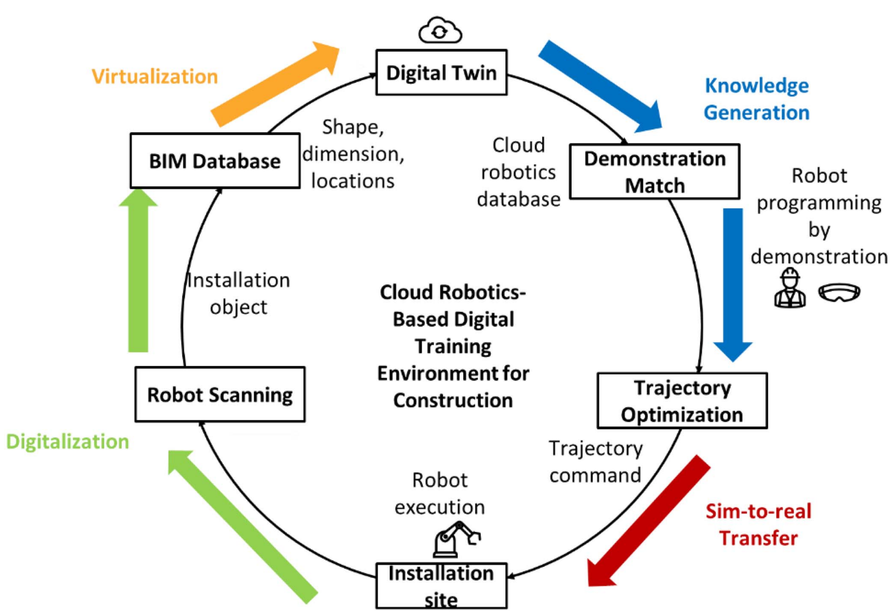

Fig. 2. Proposed workflow for cloud-robotics-based construction learning scheme.

object locations or geometric configurations, are also downloaded and used in the imitation learning database on the edge node. This process equips the robot with the capability to utilize the gathered information effectively. A hierarchical construction skill modeling system composed of generative and probability models is then used to replicate the craft skills. Subsequently, accounting for the type of robot, the edge node employs inverse kinematics (IK) (Tolani et al. 2000). Robots of various kinds can then calculate the intended robot joint states and execute the learned trajectory efficiently (Nakanishi et al. 2020). Next, the edge node utilizes wireless communication methods for real-time robot control based on the IK-calculated joint states.

Immersive Digital Training Environment and Demonstration Recording

The proposed technical contribution starts with a digital robot training environment (the DTEC), in which workers can teach by performing construction tasks naturally. As mentioned previously, the repetitive manipulation of heavy construction materials, even for demonstration purposes, causes high physical strain for construction workers. The proposed framework involves the performance of construction tasks in a virtual environment and eliminates repetitive physical interactions with heavy construction materials. The virtual environment has three components:

1. natural task demonstration capture: Motion capture starts with a hand motion tracking system with an High-Tech Computer Corporation (HTC, Taoyuan City, Taiwan), VIVE 3.0 (tracker) motion sensor, which captures human hand motions continuously to seize the demonstrated trajectories;
2. trajectory recording: The recorded trajectory is sent to the ROS, where a Linux shells script automatically launces the Rosbag file recorder and gives sound notifications on the different stages of demonstration collection. The Rosbag files are also easily connected with cloud database with AWS S3 packages.
3. digital twin: This can be readily synchronized with real-time on-site conditions using computer vision approaches, such as AprilTag-based localization and scan-to-BIM digital reconstruction (Kim and Cho 2017; Feng et al. 2021; Wang et al. 2023b); and
4. VR: A high-fidelity VR environment accommodates the digital twin and human-demonstrated motions.

The technical details of each component are described in the following.

To create an immersive and site-synchronized VR environment, we leveraged a start-of-the-art system (Wang et al. 2023b) to display target object and installation object locations, textures, and statuses in VR with low latency and high fidelity. On the basis of this VR system, box colliders were added to the digital twin objects with dimensions 1 cm larger than the original model. The colliders stop any potential collisions at a distance of 1 cm. Low-quality demonstrations with collisions can, therefore, be easily identified and avoided.

In addition, a hand motion tracking system was added. The hand tracking system was based on a base station, dongle, and tracker, as shown in Fig. 3. The base station emits infrared lasers, and sensors on the tracker receive these signals for localization. Because the localization mechanisms are nontransparent (a black box) (Bauer et al. 2021), experimental testing was used to validate localization accuracies, as shown in the experimental case study section.

To reflect human hand motions, the motion sensor was attached to the palm of a human hand. When workers move their hands in real-world demonstrations, the tracker's pose in VR will change correspondingly to reflect the construction material pose change in this demonstrated installation process. The manipulation objects in the VR environment were set as child objects of the tracker. With this setting, any motion captured by the tracker is seamlessly synchronized to the object. Therefore, hand motions for manipulating objects can be replicated accurately in VR. Moreover, to ensure applicability across different robot types, in this paper we propose the use of object trajectories as demonstrations. In the demonstration process, the VR system captures the ideal and expected object trajectories required to accomplish construction tasks. Using IK (the Kinematics and Dynamics Library, the default IK calculator for MoveIt version 1 Melodic), the desired joint states necessary to execute the ideal trajectory can be calculated. The following parameters were used for the IK motion planner (Coleman et al. 2014):

- kinematics_solver_search_resolution: 0.005
- kinematics_solver_timeout: 5
- kinematics_solver_attempts: 100

Some other IK libraries (including IK fast, Klampt, and ikpy) and different parameter settings were also tested. However, it was observed that performance rarely improved compared to the aforementioned settings.

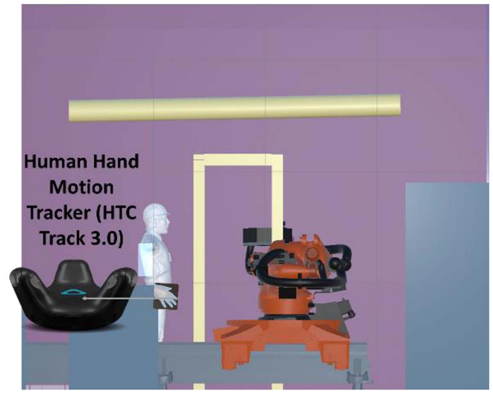

Fig. 3. Human hand motion monitoring with Tracker 3.0.

Cloud Robotics for Demonstration Storage and Information Flow

A cloud database was used to store human demonstrations and ensure convenient access without spatial and temporal limitations. The Amazon Web Services (AWS) S3 package was employed because of its ease of use. The information flow for the whole system is shown in Fig. 4. First, the demonstrations are transmitted to the ROS with ROS# protocol. Second, a Linux shell script automatically saves the trajectories as rosbag files and uses the AWS S3 functions to upload the demonstration files to the cloud. These unorganized cloud data form a cloud data lake in the AWS cloud database. When the robot needs certain knowledge or information from the cloud, it downloads the data from the cloud database using hypertext transfer protocol secure (HTTPS) protocols.

A dual storage scheme was introduced to sustainably utilize the computational output. As shown in Fig. 4, the extracted knowledge is first utilized to guide on-site robot installations, thereby enhancing the success rate of task execution. Second, the analyzed human motion patterns and knowledge are sent back to the cloud database in Python pickle models and .csv files that store the elemental motions/motion primitives for the assigned tasks. The extracted elemental motions/motion primitives can be used for future robot installation tasks and human worker education. By combining the use of extracted knowledge with the storage of motion patterns in a cloud database, the system becomes more robust and capable of handling future installation challenges effectively.

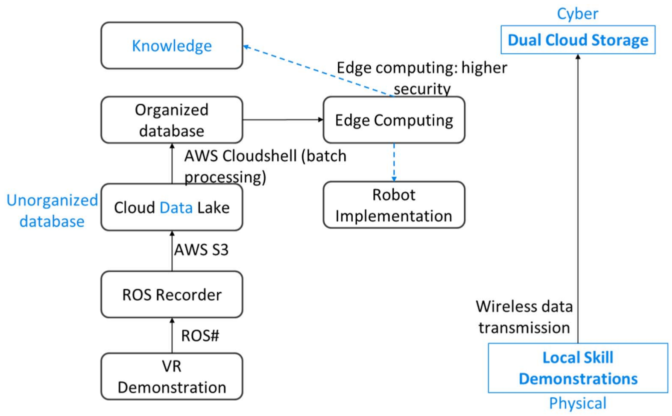

Fig. 4. Information flow in the cloud robotics framework.

Federated Robot Learning and Data Collection

With the proposed dual storage scheme, knowledge and demonstration data can be saved in the cloud database and used at any time for any task. To alleviate the burden of repetitive demonstrations, we propose the generalization of previously learned knowledge by collecting demonstration data from heterogeneous tasks, such as tasks from various locations and installation targets of diverse geometric configurations. Such diverse data contribute to the generalizability of robot learning—in contrast to traditional single-source robot learning data collection. This heterogeneous and crowdsourced demonstration data naturally forms a federated data collection scheme.

For a newly encountered task, the trajectory is transformed to the new task scenario by twisting the pose (x, y, z) with the task parameter variations, as shown in Eq. (1):

$$p' = \begin{pmatrix} \boldsymbol{x}' \\ \boldsymbol{y}' \\ \boldsymbol{z}' \\ \mathbf{1} \end{pmatrix} = T \cdot p = \begin{pmatrix} \boldsymbol{a}_{11} & \boldsymbol{a}_{12} & \boldsymbol{a}_{13} & \boldsymbol{a}_{14} \\ \boldsymbol{a}_{21} & \boldsymbol{a}_{22} & \boldsymbol{a}_{23} & \boldsymbol{a}_{24} \\ \boldsymbol{a}_{31} & \boldsymbol{a}_{32} & \boldsymbol{a}_{33} & \boldsymbol{a}_{34} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} \end{pmatrix} \cdot \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \\ \boldsymbol{z} \\ \mathbf{1} \end{pmatrix} \tag{1}$$

with T: task parameter transformation matrix

$$a_{11} := x^{\text{new}} - x^{\text{old}}; \qquad a_{22} := y^{\text{new}} - y^{\text{old}}; \qquad a_{11} := z^{\text{new}} - z^{\text{old}}$$

In this way, trajectory demonstrations from diverse task locations can be transformed to have the same target location and concatenated with the new ad hoc demonstrations to form an expanded database. The performance of the resultant robot learning model with expanded data set is illustrated in the experimental results section.

Construction Task Decomposition and Hierarchical Modeling

Once demonstration data sets are prepared, the next step is the hierarchical decomposition and learning of construction skills. In traditional construction master–apprentice relationships between human workers, the acquisition of construction skills heavily relies on muscle memory, practical experience, and hands-on work exposure (Sing et al. 2017). In the domain of robot learning and programming, one responsible and explainable approach to replicating such apprenticeship is motion primitives to decompose a complex, long-horizon task into multiple elemental motions (Schaal et al. 2005), as shown in Fig. 5.

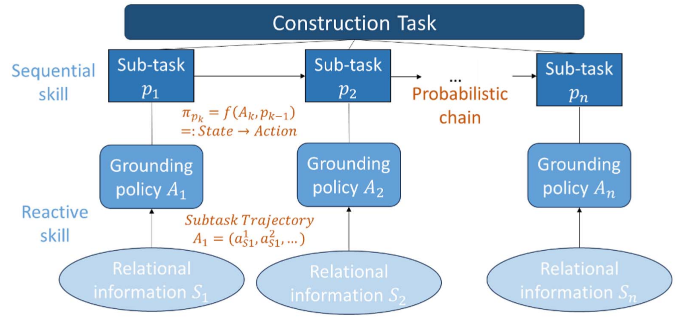

Fig. 5. Construction task decomposition.

In robot motion primitive learning from demonstrations, both predetermined elemental motions (Xie et al. 2020) and automatic extraction of elemental motions from observations (Cohen et al. 2021) have been used. We chose the predetermined approach for the following reasons. First, construction sites are generally highly unstructured (Golparvar-Fard et al. 2013; Jeong et al. 2021), and demonstration trajectories may be affected by the dynamic environment of one site. Motions extracted from such sites are not representative enough to be generalized to all other sites. Second, there is ongoing research on decomposing construction tasks into activities and smaller elemental motions [e.g., Wu et al. (2022)]. Wu et al.'s paper decomposed bricklaying activity into guiding bricks, placing bricks, and cutting bricks. Using predetermined motions provides easier and more convenient connections with high-quality construction management studies like this. The connected approach may also be more efficient than collecting a high quantity of data only to extract the elemental motions.

The first step in determining motion primitives is to find the number of elemental motions. We applied k-means clustering on the average of three demonstrated trajectories. The lowest cluster error was $k = 8$, so the number of elemental motions was set to eight. Collaborative work with an experienced worker was initiated to develop the name and meaning of each elemental motion based on the spatial characteristic of each step—that is, the average coordinates of each elemental motion. The resultant eight elemental steps of ceiling installation are shown in Fig. 6. The results of this clustering algorithm are still affected by the personal habits of demonstrators. Considering that the main scope of this paper is to propose a feasible cyber VR learning and knowledge generation architecture, we are exploring construction segmentation and clustering optimizations in our ongoing work.

As shown in Fig. 6, each motion primitive represents an atomic motion or subtask that forms a fundamental unit of the overall task.

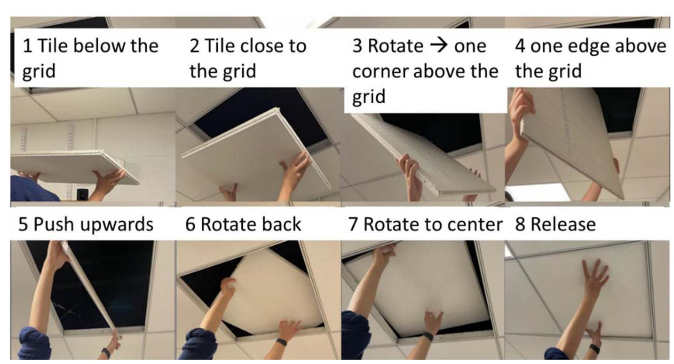

Fig. 6. Motion primitives for ceiling installation tasks.

The entire task is treated as a probabilistic chain that connects each step of the motion primitives. Each motion primitive also reflects the human worker's reaction to relational information observed from the environment, such as the distance between material that has been picked up to the target location. To categorize this decision-making process more comprehensively, we propose hierarchical construction skill models comprising sequential and reactive skills. The details of these two skills are described in the following.

Sequential Skills

Sequential skills are proposed to avoid perceptual aliasing, which conditions the choice of actions to the previous history of actions, and avoid confusion with similar states observed for different subtasks. With previous actions forming the conditional probability, the probability of mislabeling environmental observations to those in other subtasks is largely eliminated.

However, the preliminary condition of the probability chain is that the robot agent remembers the history of actions, which implies the necessity of memory capability in the corresponding computation model. A gated recurrent unit (GRU) suits this purpose—it has the capability to store the history of past actions and use it to update the next action decisions. The eight elemental motions are treated as time-series data and used as the input of the GRU model. As shown in Fig. 7, the GRU's gates use both the previous hidden state and the current input (the actions taken in the previous state) to decide the output of actions. Therefore, it forms the elemental chain and demonstrates sequential skills by deciding future actions based on the history of actions.

In addition, we tested some other models that also condition the current action based on previous actions, such as the hidden Markov model (HMM) for high-level sequential skill modeling. Because the GRU model significantly outperformed the HMM in terms of learning performance, this paper only illustrates task decomposition with the GRU model for sequential skills.

At each update and reset gate, the input (the observed last action) and hidden state are passed to the activation function to codetermine the next step, as follows:

$$R_t = \sigma(X_{t-1}W_{t1} + H_{t-1}W_{t2} + b_r)$$

$$U_t = \sigma(X_{t-1}W_{t3} + H_{t-1}W_{t4} + b_u)$$

where R_t and U_t = reset and update gates at time t, respectively; X_{t-1} and H_{t-1} = input data of previous action histories and hidden variables, respectively; and W_{t1}, W_{t2}, W_{t3}, W_{t4}, b_r, and b_u = weight parameters.

Reactive Skills

Reactive skills represent how human actions are influenced by environmental observations. The observed object state (A) is mapped to a motion primitive (S) through a generative variational autoencoder (VAE) (Kingma and Welling 2013) model, $A = f(S)$. The VAE function f is shown in Fig. 8.

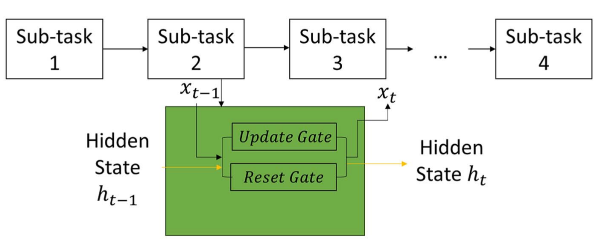

Fig. 7. High-level construction task decomposition with sequential skills.

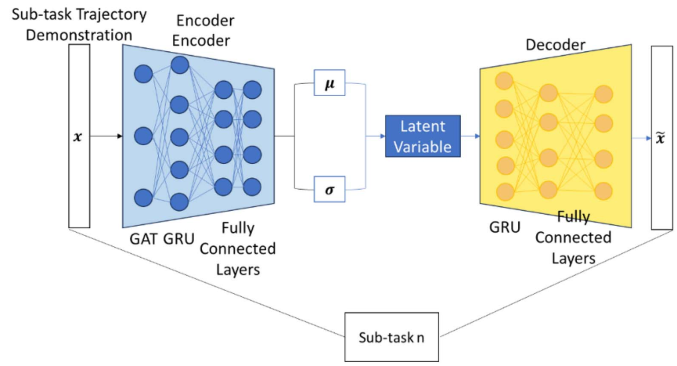

Fig. 8. Low-level reactive skill to ground subtask action policies.

The VAE model in Fig. 8 has the following structure and training parameters:

- Model structure: encoder and decoder, as shown in Fig. 8
- Encoder model: GAT-GRU-FC*7-Softmax (Xie et al. 2020)
- Decoder model: GRU-FC*7-Softmax (Xie et al. 2020)
- Latent size: 512
- Training epochs: 1,000
- Optimizer: Adam (Kingma and Ba 2014)
- Learning rate (both encoder and decoder): 0.0001

The model training process was finished with PyTorch (Paszke et al. 2019).

The input for the lower level of subtask learning is composed of two parts. The first part is the task parameters. For the case study on ceiling installations, we chose the model dimensions and the location of the ceiling grid to connect with the task information stored in the cloud BIM database. This information was used to preprocess and prepare the demonstration data sets. The second part of the data input for the VAE model is the installation object state, which includes the 6 degree of freedom (DOF) spatial motions of three-dimensional (3D) linear motions (x, y, z) and 3D rotations (r, p, y). We used the installation object and the goal pose—the absolute pose of the ceiling grid. Another option is to use the object's relative pose (Zha et al. 2021; Luo et al. 2023). Because the first option has achieved satisfactory results and the goal of this study was to provide a working system, we did not compared the two options.

Evaluation Case Study

Ceiling Installation Skill Modeling and Transfer

A case study was conducted to evaluate the proposed approach to transferring ceiling installation skills to the robot. As described in Liang et al. (2020), ceiling installation is a construction task with one of the highest requirements for dexterity and environmental adaptations. Ceiling tiles are generally larger than the slots in the ceiling grid and need some manipulation in order to be suspended on a grid. Most motion planning algorithms focus on avoiding obstacles rather than maneuvering through them; therefore, they cannot be used for ceiling installation tasks (Liang et al. 2020; Yang et al. 2023).

By observing one experienced construction worker practice the ceiling installation task, eight elemental motions were summarized, as shown in Fig. 6. A human subject observed and replicated the installation in VR for repetitive demonstration data collection. In the next step, eight VAE models were trained for each elemental motion to represent the reactive skills. A GRU model connected them and showed the sequential skills (Xie et al. 2020).

The object was a standard two-by-two ceiling grid, which used the most common ceiling tile size. The upper left point of the grid had coordinates of (2.95, 1.5, 2). The eight elemental motions in Fig. 6 were used to form a continuous demonstration. However, these elemental motions were separately recorded in eight different rosbag files. The recording of the demonstration for each elemental motion lasted for 5 s (a tested threshold to ensure the smooth completion of each motion), and the demonstrator was notified with a beep every 5 s. Five demonstrations were collected in the DTEC to train the robot to perform this task (Data Set 1). To expand the database, ten more demonstrations were performed by the same person and stored in the cloud database as follows: The 5 demonstrations are referring to the number of new demonstrations and 10 is referring to the total number of demonstrations in Data Set 2. Likewise for Data Set 3. To avoid confusions, the authors propose to change the parenthetical statements to: (adding 5 more demonstrations to Data Set 1 to form Data Set 2. Data Set 2 has a total of 10 demonstrations) and (adding 5 more demonstrations to Data Set 2 to form Data Set 3. Data Set 3 has a total of 15 demonstrations). The number of data points collected varied from 382 to more than 400 due to the data transmission speed. To avoid an imbalance in the training data, the data for each episode of the elemental motion was truncated to 380 points. The computational results are shown in the results section.

A KUKA KR120 robot (Augsburg, Germany), a 6DOF robot with a workload of 120 kg, was used to execute the task. In the first several batches of simulation robot experiments, the robot exhibited some jitter. The jittery behavior was caused by excessively small increments in the demonstrated and learned trajectories. To solve this issue, an adaptive window filter with a window size of 0.1 was adopted to smooth the trajectories; the filter smooths the trajectory until the difference between the current or averaged coordinate and the next waypoint exceeds the given threshold.

Simulation of Real World Skill Transfer and Robot Motion Control

With the demonstration data collected, stored, and analyzed in the previous steps, the next step involved determining how to use these data for robot motion control. One major obstacle was coordinate misalignment between the DTEC environment and the robot control environment. The coordinate transformation matrix was calculated with Algorithm 1, shown in the following. A total of 728 readings from five locations (the four corners and centroid of the ceiling grid) were used. The average localization error was 0.0019 m, including 3DOF location and 4DOF quaternion orientation.

Algorithm 1. DTEC-gazebo coordinates transfer

For each Rosbag file i=0,1,2,3,4:
  extract the Unity pose information for each stamp
  $p^{ut} = [x^{ut}, y^{ut}, z^{ut}, ox^{ut}, oy^{ut}, oz^{ut}, ow^{ut}]$
  remove outliers
  add to a vector for batch processing
  For all extracted poses Unity p^{ut}:
    calculate real world pose: $p^{wt} = [x^{ut}, y^{ut}, z^{ut}, 0, 0, 0, 1]$
    calculate transformation matrix T^t: $T^t = p^{wt-1} \cdot p^{ut}$
    add T^t to a vector and return its mean value of T^i
return $T^{cal} = mean(T^0 \sim T^4)$
For all p^{ut}:
  calculate $p^{wt'} = T^{cal} \cdot p^{ut}$
  calculate error: $e = p^{wt'} - p^{wt}$
return mean error

Experimental Results and Discussion

The proposed cloud robotics framework and hierarchical robot learning model were evaluated for improvements in skill transfer. The following evaluation metrics were adopted: (1) learning model training time and error with different demonstration sources; and (2) robot task execution success rate with different demonstration sources.

Learning Model Performance

Model training time and mean square error (MSE) loss were used as evaluation metrics based on established practices from federated learning research and robot HIL algorithms (Lin et al. 2022; Yang et al. 2020; Xie et al. 2020; Hayashi et al. 2022). Model training time shows how long it takes for a learning model to achieve convergence. In this study, variations in training speed reflected data set quality, because a consistent model architecture was employed (Yang et al. 2020). They also offered insights into the efficacy of introducing heterogeneous task demonstrations to enhance the quality of the cloud database. Second, MSE loss is widely used for unsupervised learning models in which the outputs comprise predictions or trajectories. It shows the average euclidean distance between the demonstration and the learning-model-generated trajectory (Xie et al. 2020; Lin et al. 2022). In this case, the sequential skill learning model returned zero training and validation errors. The computation results for reactive skill modeling, in terms of the reduced learning errors and increased learning speed, are shown in Fig. 9 and Table 1. The training was conducted on a computer with a GeForce RTX 3060 graphics processing unit (GPU) (NVIDIA, Santa Clara, CA). The numerical results of the computational errors and learning speed vary depending on the computers used.

Fig. 9 and Table 1 show that for each epoch, the Data Set 1 (the data set with only one installation location) often had the largest error. This shows that when the model learned from Data Set 1, it generated a trajectory that was the most different from the demonstrations, which did not represent the human-demonstrated skills. This was due to the small amount and homogeneity of the data contained in this data set (only five demonstrations of a two-by-two ceiling tile installation in one location). Regarding the two other data sets: (1) because they exist in the cloud database already,

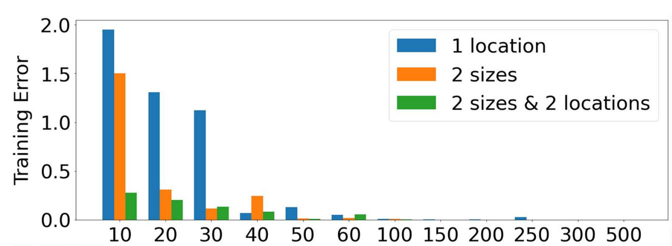

Fig. 9. Training error comparison for three data sets: 10–500 epochs.

Table 1. Mean and variance of model learning speed and errors

Data Set	Training time		Training error	
	Mean	Variance	Mean	Variance
Data Set 1: one size and one location	19.356	405.142	0.389	0.454
Data Set 2: two sizes and one location	19.264	399.528	0.184	0.183
Data Set 3: two sizes and two locations	19.115	389.370	0.0639	0.00885

J. Comput. Civ. Eng., 2024, 38(4): 04024019

Fig. 10. A KUKA KR120 robot following the demonstrated motion primitives.

the workers do not need to make an extra effort to generate them again; and (2) with an increase of heterogeneous task demonstrations, model learning error is largely reduced, especially with fewer training epochs.

Moreover, as the statistical features of the data sets in Table 1 show, Data Set 3 had the lowest mean value of training time and training error, illustrating the potent learning power provided by the heterogeneous data. In addition, the variance of the training error and time from Data Set 3 was the lowest. Because the variance shows how the number of training epochs affects model performance, the small number suggests that using this data set helped the model converge early (i.e., converge with fewer epochs) and be more robust for noisy data.

Robot Task Execution Performance

Fig. 10 shows the physical robot following the given motion primitives.

The evaluation framework aligned with the metrics used by Liang et al. (2020); a 5-mm allowance between a tile's ultimate location and the grid was permissible. Fig. 11 shows a successful installation instance (on the left) and a case that exceeded the threshold (on the right).

The batch experiment of robot execution was performed in Gazebo version 9, in order to avoid damaging the workpiece. The experiment was designed to evaluate robot task performance for ceiling installation in different locations with demonstration data

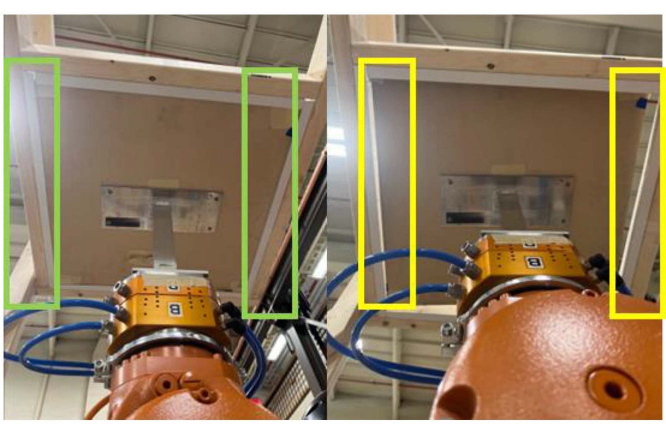

Fig. 11. Robot task execution evaluation criteria example.

from one or both target locations. Data Sets 1 and 3 were chosen to train the skill learning model, mainly because Data Set 2 was the same as Data Set 1 in this case. The results of these evaluations is shown in Table 2, which enumerates the success rates across discrete experimental scenarios with five repetitions for each scenario. Table 2 shows that when the robot only used the five ad hoc demonstrations, it successfully finished the given tasks 80% of the time. However, the robot cannot generalize this knowledge to new task scenarios with a different target destination—it collides with the grid and does not move to the designated target location—even with another round of training, especially for new locations. Yet, with the same training process (one for the original location and another for the new location), this problem is well addressed with five more demonstrations in new task scenarios in Data Set 3. The KUKA robot is capable of finishing installation tasks at both (2.95, 1.5, 2) and (3.5, 2, 2).

Limitations

Although we accomplished our research goals, complete with experimental validations, we acknowledge that there were several limitations to this study.

First, although this paper addressed task heterogeneity, the demonstration data still possess a certain level of homogeneity, because the demonstrations were performed by only one person. It was assumed that the existing data in the cloud were also from this person. The literature (Chen et al. 2015; He 2017; de Luca et al. 2022; Mendieta et al. 2022) shows that having data from different people may improve or reduce learning performance. However, performance drops can still be managed through data preprocessing approaches, joint modeling, or unifying algorithms (Mandal et al. 2019; Paleja and Gombolay 2019; Nasiriany et al. 2022; Kuhar et al. 2023). In future studies, the cloud database architecture and corresponding data selection, augmentation, and unifying methods will be explored.

Second, this paper addressed skill transfers from humans to robots, describing a case study that illustrated the proposed framework with a ceiling tile installation and ceiling grid avoidance skill transfers. However, there are many other challenges on a highly dynamic and unstructured construction site that were not addressed in our case study. Future studies will explore the applicability of the proposed approach in transferring other human skills to robots. For example, there may be many utility pipes around a ceiling grid. Utility pipe and ceiling grid coavoidance skills require high

Table 2. Simulated KUKA robot task execution performance comparisons

Data Set	Original location (%)	New location (%)	New location failure reasons
Data Set 1: one location	80	0	Robot collided with ceiling grid
Data Set 3: two locations	100	80	Unable to find valid motion plans

environmental adaptation, flexibility, and dexterity. In the future, we will explore transferring these skills with the current or an improved version of the cloud framework.

Third, the proposed construction task decomposition approach and subtask hierarchies are only demonstrating one feasible way to decode the construction tasks. There are different ways to decompose a single task. For example, motion primitives are currently defined based on the pose, including location and orientation, of task objects. Future work can explore construction task decomposition variations to optimize the proposed hierarchical representation of construction installation tasks. One committed goal of this work is to reduce the workload of human workers.

Fourth, this current work only forms an open-loop control system of the construction robots. Data preprocessing, uploading to the cloud, downloading to local devices, and edge learning are all in separate scripts and need to be connected by the manual launch of several Linux scripts. Although the proposed system can be easily applied to any human worker in VR or any robot on a construction site for scalable applications, the information exchange system is not yet in real time. Future work will focus on connecting the innovations in data exchange and communication in federated learning to upgrade this system to an advanced closed-loop construction robot learning and control scheme.

Conclusions

This study targeted several problems in construction robot application and craft skill programming. Building on previous research exploring the possibility of IL to transfer craft skills from human workers to robots, this work addressed the following problems:

1. The high physical workload in repetitive physical demonstrations: The high workload is caused by (1) the manipulation of heavy construction materials; and (2) the need for repetitive demonstrations for new tasks. To address the first reason, we proposed demonstrating in a high-fidelity VR and digital twin environment to avoid interacting with heavy materials. To address the second reason, we developed a cloud database to connect to VR and ROS to save past demonstrations.
2. The low generalizability in robot IL: The majority of robot IL models, including RL-based ones, depend heavily on environmental observations and task parameters. In this study, we proposed a data-driven approach with a federated demonstration collection scheme and saved crowdsourced data in a cloud database. With more variation in the learning data, the learning model naturally possesses generalizable skills. Experiments conducted with a simulated robot also validated this conclusion. For example, as Table 2 shows, the robot cannot generalize the knowledge learned from one location to tasks in new locations with data from a single location. However, with data from two locations, the skill learned from Data Set 3 is generalizable. In addition, because we purposefully designed the demonstration to be manipulated workpiece trajectories, learned trajectories can be generated for different robots, such as the KUKA KR120 prototyping robot.
3. The lack of learning model transparency and corresponding reduced trustworthiness: Traditional RL-based structure learns the craft skills in a black box. Human users—especially workers without programming expertise—find it challenging to understand and trust robot learning models (Gunning 2017). The proposed explainable hierarchical model decodes craft skills as high-level sequential skills of subtasks and low-level reactive skills that map environmental observations to a certain choice of subtask. A generative VAE model was used to model reactive skills with a very low average error of 0.000895 m with 500 epochs of training; the GRU model determined the sequence of actions with a mean error of 0. Because both models achieved satisfactory performance, the craft skills were also decoded and recorded in interpretable robot learning models.

To summarize, in this paper we proposed a scalable scheme to program construction robots with craft skills. The proposed approach features cloud robotics and hierarchical learning to reduce the workload on workers and the need for repetitive demonstrations. The approach's enhanced explainability is also more user-friendly for workers without programming expertise.

Data Availability Statement

Some data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request, including the ceiling installation demonstration trajectory data.

Acknowledgments

The authors would like to acknowledge the financial support for this research received from the US National Science Foundation (NSF) (Grant Nos. FW-HTF 2025805 and FW-HTF 2128623). Any opinions and findings in this paper are those of the authors and do not necessarily represent those of the NSF. The authors would also like to acknowledge the contribution of engineering technician Justin Roelofs for demonstrating the ceiling installation process and building the wooden models used in the physical robot experiments.

References

Abdo, N., H. Kretzschmar, and C. Stachniss. 2012. "From low-level trajectory demonstrations to symbolic actions for planning." In *Proc., ICAPS Workshop on Combining Task and Motion Planning for Real-World App*, 29–36. State College, PA: Citeseer.

Adami, P., P. B. Rodrigues, P. J. Woods, B. Becerik-Gerber, L. Soibelman, Y. Copur-Gencturk, and G. Lucas. 2022. "Impact of VR-based training on human–robot interaction for remote operating construction robots." *J. Comput. Civ. Eng.* 36 (3): 04022006. https://doi.org/10.1061/(ASCE)CP.1943-5487.0001016.

AGC (Associated General Contractors). 2016. *Two-thirds of contractors have a hard time finding qualified craft workers to hire amid growing construction demand, national survey finds*. Arlington, VA: AGC.

Aryan, A., F. Bosché, and P. Tang. 2021. "Planning for terrestrial laser scanning in construction: A review." *Autom. Constr.* 125 (Mar): 103551. https://doi.org/10.1016/j.autcon.2021.103551.

Balaji, B., et al. 2016. "Brick: Towards a unified metadata schema for buildings." In *Proc., 3rd ACM Int. Conf. on Systems for Energy-Efficient*

Built Environments, 41–50. New York: Association for Computing Machinery.

Bandera, J. P., L. Molina-Tanco, J. A. Rodriguez, and A. Bandera. 2010. Architecture for a robot learning by imitation system. In *Proc., Melecon 2010-2010 15th IEEE Mediterranean Electrotechnical Conf.*, 87–92. New York: IEEE.

Bauer, P., W. Lienhart, and S. Jost. 2021. "Accuracy investigation of the pose determination of a VR system." *Sensors* 21 (5): 1622. https://doi.org/10.3390/s21051622.

Bello, S. A., L. O. Oyedele, O. O. Akinade, M. Bilal, J. M. D. Delgado, L. A. Akanbi, A. O. Ajayi, and H. A. Owolabi. 2021. "Cloud computing in construction industry: Use cases, benefits and challenges." *Autom. Constr.* 122 (Feb): 103441. https://doi.org/10.1016/j.autcon.2020.103441.

BLS (US Bureau of Labor). 2023. "Industries at a glance, construction: NAICS 23." Accessed June 16, 2023. https://www.bls.gov/iag/tgs/iag23.htm.

Brosque, C., and M. Fischer. 2022. "Safety, quality, schedule, and cost impacts of ten construction robots." *Constr. Rob.* 6 (2): 163–186. https://doi.org/10.1007/s41693-022-00072-5.

Cai, J., A. Du, X. Liang, and S. Li. 2023. "Prediction-based path planning for safe and efficient human–robot collaboration in construction via deep reinforcement learning." *J. Comput. Civ. Eng.* 37 (1): 04022046. https://doi.org/10.1061/(ASCE)CP.1943-5487.0001056.

Chen, Z. Y., Z. P. Fan, and M. Sun. 2015. "Behavior-aware user response modeling in social media: Learning from diverse heterogeneous data." *Eur. J. Oper. Res.* 241 (2): 422–434. https://doi.org/10.1016/j.ejor.2014.09.008.

Cohen, Y., O. Bar-Shira, and S. Berman. 2021. "Motion adaptation based on learning the manifold of task and dynamic movement primitive parameters." *Robotica* 39 (7): 1299–1315. https://doi.org/10.1017/S0263574720001186.

Coleman, D., I. Sucan, S. Chitta, and N. Correll. 2014. "Reducing the barrier to entry of complex robotic software: A Moveit! case study." Preprint, submitted April 15, 2014. http://arxiv.org/abs/1404.3785.

DelPreto, J., J. I. Lipton, L. Sanneman, A. J. Fay, C. Fourie, C. Choi, and D. Rus. 2020. "Helping robots learn: A human-robot master-apprentice model using demonstrations via virtual reality teleoperation." In *Proc., 2020 IEEE Int. Conf. on Robotics and Automation (ICRA)*, 10226–10233. New York: IEEE.

de Luca, A. B., G. Zhang, X. Chen, and Y. Yu. 2022. "Mitigating data heterogeneity in federated learning with data augmentation." Preprint, submitted June 20, 2022. http://arxiv.org/abs/2206.09979.

Delvinne, H. H., K. Hurtado, J. Smithwick, B. Lines, and K. Sullivan. 2020. "Construction workforce challenges and solutions: A national study of the roofing sector in the United States." In *Proc., Construction Research Congress 2020: Safety, Workforce, and Education*, 529–537. Reston, VA: ASCE.

Deng, M., B. Fu, C. C. Menassa, and V. R. Kamat. 2023. "Learning-based personal models for joint optimization of thermal comfort and energy consumption in flexible workplaces." *Energy Build.* 298 (Jun): 113438. https://doi.org/10.1016/j.enbuild.2023.113438.

Du, J., Y. Shi, Z. Zou, and D. Zhao. 2018. "CoVR: Cloud-based multiuser virtual reality headset system for project communication of remote users." *J. Constr. Eng. Manage.* 144 (2): 04017109. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001426.

Fang, Y., and Y. K. Cho. 2016. "Real-time visualization of crane lifting operation in virtual reality." In Vol. 11 of *Proc., 16th Int. Conf. on Construction Applications of Virtual Reality*, 13. Hong Kong: Hong Kong Univ. of Science and Technology.

Feng, C., V. R. Kamat, and C. C. Menassa. 2016. "Marker-assisted structure from motion for 3D environment modeling and object pose estimation." In *Proc., Construction Research Congress 2016*, 2604–2613. Reston, VA: ASCE.

Feng, Y., J. Wang, H. Fan, and Y. Hu. 2021. "A BIM-based coordination support system for emergency response." *IEEE Access* 9 (Jun): 68814–68825. https://doi.org/10.1109/ACCESS.2021.3077237.

Golparvar-Fard, M., A. Heydarian, and J. C. Niebles. 2013. "Vision-based action recognition of earthmoving equipment using spatio-temporal features and support vector machine classifiers." *Adv. Eng. Inf.* 27 (4): 652–663. https://doi.org/10.1016/j.aei.2013.09.001.

Gunning, D. 2017. "Explainable artificial intelligence (xai), 2017." In *Defense advanced research projects agency (DARPA) project*. Ft. Belvoir, VA: Defense Technical Information Center.

Haddadin, S., A. De Luca, and A. Albu-Schäffer. 2017. "Robot collisions: A survey on detection, isolation, and identification." *IEEE Trans. Rob.* 33 (6): 1292–1312. https://doi.org/10.1109/TRO.2017.2723903.

Hamon, R., H. Junklewitz, and I. Sanchez. 2020. *Robustness and explainability of artificial intelligence—From technical to policy solutions, EUR 30040*. Luxembourg: Publications Office of the European Union.

Hayashi, K., S. Sakaino, and T. Tsuji. 2022. "An independently learnable hierarchical model for bilateral control-based imitation learning applications." *IEEE Access* 10 (Apr): 32766–32781. https://doi.org/10.1109/ACCESS.2022.3155255.

He, J. 2017. "Learning from data heterogeneity: Algorithms and applications." In *Proc., IJCAI*, 5126–5130. San Francisco: International Joint Conferences on Artificial Intelligence.

Hsiao, K., and T. Lozano-Perez. 2006 "Imitation learning of whole-body grasps." In *Proc., 2006 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 5657–5662. New York: IEEE.

Huang, L., Z. Zhu, and Z. Zou. 2023. "To imitate or not to imitate: Boosting reinforcement learning-based construction robotic control for long-horizon tasks using virtual demonstrations." *Autom. Constr.* 146 (Feb): 104691. https://doi.org/10.1016/j.autcon.2022.104691.

Jeong, I., Y. Jang, J. Park, and Y. K. Cho. 2021. "Motion planning of mobile robots for autonomous navigation on uneven ground surfaces." *J. Comput. Civ. Eng.* 35 (3): 04021001. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000963.

Jiao, Y., Y. Wang, S. Zhang, Y. Li, B. Yang, and L. Yuan. 2013. "A cloud approach to unified lifecycle data management in architecture, engineering, construction and facilities management: Integrating BIMs and SNS." *Adv. Eng. Inf.* 27 (2): 173–188. https://doi.org/10.1016/j.aei.2012.11.006.

Karp, P. D., S. M. Paley, and I. Greenberg. 1994. "A storage system for scalable knowledge representation." In *Proc., Third Int. Conf. on Information and Knowledge Management*, 97–104. New York: Association for Computing Machinery.

Kase, K., C. Paxton, H. Mazhar, T. Ogata, and D. Fox. 2020. "Transferable task execution from pixels through deep planning domain learning." In *Proc., 2020 IEEE Int. Conf. on Robotics and Automation (ICRA)*, 10459–10465. New York: IEEE.

Kim, P., and Y. K. Cho. 2017. "An automatic robust point cloud registration on construction sites." In *Computing in civil engineering 2017*, 411–419. Reston, VA: ASCE.

Kingma, D. P., and J. Ba. 2014. "Adam: A method for stochastic optimization." Preprint, submitted December 22, 2014. https://arxiv.org/abs/1412.6980.

Kingma, D. P., and M. Welling. 2013. "Auto-encoding variational bayes." In *Proc., 2nd Int. Conf. on Learning Representations (ICLR)*. Amherst, MA: Univ. of Massachusetts Amherst.

Kohler, M. D., R. W. Clayton, Y. Bozorgnia, E. Taciroglu, and R. Guy. 2022. "The community seismic network: Applications and expansion to 1200 stations." In Vol. 2022 of *Proc., AGU Fall Meeting Abstracts*. Washington, DC: American Geophysical Union.

Kuhar, S., S. Cheng, S. Chopra, M. Bronars, and D. Xu. 2023. "Learning to discern: Imitating heterogeneous human demonstrations with preference and representation learning." In *Proc., Conf. on Robot Learning*, 1437–1449. Cambridge, MA: Proceedings of Machine Learning Research.

Kyrarini, M., M. A. Haseeb, D. Ristić-Durrant, and A. Gräser. 2019. "Robot learning of industrial assembly task via human demonstrations." *Auton. Robots* 43 (1): 239–257. https://doi.org/10.1007/s10514-018-9725-6.

Lee, D., S. Lee, N. Masoud, M. S. Krishnan, and V. C. Li. 2022. "Digital twin-driven deep reinforcement learning for adaptive task allocation in robotic construction." *Adv. Eng. Inf.* 53 (Aug): 101710. https://doi.org/10.1016/j.aei.2022.101710.

Liang, C. J., and M. H. Cheng. 2023. "Trends in robotics research in occupational safety and health: A scientometric analysis and review." *Int. J. Environ. Res. Public Health* 20 (10): 5904. https://doi.org/10.3390/ijerph20105904.

Liang, C. J., V. R. Kamat, and C. C. Menassa. 2020. "Teaching robots to perform quasi-repetitive construction tasks through human demonstration." *Autom. Constr.* 120 (Dec): 103370. https://doi.org/10.1016/j.autcon.2020.103370.

Liang, C. J., T. H. Le, Y. Ham, B. R. Mantha, M. H. Cheng, and J. J. Lin. 2023. "Ethics of artificial intelligence and robotics in the architecture, engineering, and construction industry." Preprint, submitted October 9, 2023. https://arxiv.org/abs/2310.05414.

Liang, C. J., X. Wang, V. R. Kamat, and C. C. Menassa. 2021. "Human–robot collaboration in construction: Classification and research trends." *J. Constr. Eng. Manage.* 147 (10): 03121006. https://doi.org/10.1061/(ASCE)CO.1943-7862.0002154.

Lin, X., Z. Huang, Y. Li, J. B. Tenenbaum, D. Held, and C. Gan. 2022. "Diffskill: Skill abstraction from differentiable physics for deformable object manipulations with tools." Preprint, submitted March 21, 2022. http://arxiv.org/abs/2203.17275.

Liu, C., S. Shirowzhan, S. M. E. Sepasgozar, and A. Kaboli. 2019. "Evaluation of classical operators and fuzzy logic algorithms for edge detection of panels at exterior cladding of buildings." *Buildings* 9 (2): 40. https://doi.org/10.3390/buildings9020040.

Liu, H., S. Nasiriany, L. Zhang, Z. Bao, and Y. Zhu. 2022. "Robot learning on the job: Human-in-the-loop autonomy and learning during deployment." Preprint, submitted July 4, 2023. http://arxiv.org/abs/2211.08416.

Lundeen, K., V. Kamat, C. Menassa, and W. McGee. 2018. "Adaptive perception and modeling for robotized construction joint filling." In *Proc., Int. Symp. on Automation and Robotics in Construction (ISARC)*, 244–251. Oulu, Finland: International Association for Automation and Robotics in Construction.

Lundeen, K. M., V. R. Kamat, C. C. Menassa, and W. McGee. 2017. "Scene understanding for adaptive manipulation in robotized construction work." *Autom. Constr.* 82 (Oct): 16–30. https://doi.org/10.1016/j.autcon.2017.06.022.

Luo, J., C. Xu, X. Geng, G. Feng, K. Fang, L. Tan, S. Schall, and S. Levine. 2023. "Multi-stage cable routing through hierarchical imitation learning." *IEEE Trans. Rob.* 40 (Jan): 1476–1491. https://doi.org/10.1109/TRO.2024.3353075.

Makondo, N., B. Rosman, and O. Hasegawa. 2015. "Knowledge transfer for learning robot models via local Procrustes analysis." In *Proc., 2015 IEEE-RAS 15th Int. Conf. on Humanoid Robots (Humanoids)*, 1075–1082. New York: IEEE.

Mandal, S. K., G. Bhat, C. A. Patil, J. R. Doppa, P. P. Pande, and U. Y. Ogras. 2019. "Dynamic resource management of heterogeneous mobile platforms via imitation learning." *IEEE Trans. Very Large Scale Integr. VLSI Syst.* 27 (12): 2842–2854. https://doi.org/10.1109/TVLSI.2019.2926106.

Mendieta, M., T. Yang, P. Wang, M. Lee, Z. Ding, and C. Chen. 2022. "Local learning matters: Rethinking data heterogeneity in federated learning." In *Proc., IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, 8397–8406. New York: IEEE.

Nakanishi, J., S. Itadera, T. Aoyama, and Y. Hasegawa. 2020. "Towards the development of an intuitive teleoperation system for human support robot using a VR device." *Adv. Rob.* 34 (19): 1239–1253. https://doi.org/10.1080/01691864.2020.1813623.

Nasiriany, S., T. Gao, A. Mandlekar, and Y. Zhu. 2022. "Learning and retrieval from prior data for skill-based imitation learning." Preprint, submitted November 15, 2022. http://arxiv.org/abs/2210.11435.

Neto, P., and N. Mendes. 2013. "Direct off-line robot programming via a common CAD package." *Rob. Auton. Syst.* 61 (8): 896–910. https://doi.org/10.1016/j.robot.2013.02.005.

Paleja, R., and M. Gombolay. 2019. "Heterogeneous learning from demonstration." In *Proc., 14th ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI)*, 730–732. New York: IEEE.

Pan, M., and W. Pan. 2020. "Stakeholder perceptions of the future application of construction robots for buildings in a dialectical system framework." *J. Manage. Eng.* 36 (6): 04020080. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000846.

Park, S., H. Yu, C. C. Menassa, and V. R. Kamat. 2023. "A comprehensive evaluation of factors influencing acceptance of robotic assistants in field construction work." *J. Manage. Eng.* 39 (3): 04023010. https://doi.org/10.1061/JMENEA.MEENG-5227.

Paszke, A., et al. 2019. "Pytorch: An imperative style, high-performance deep learning library." In *Advances in neural information processing systems*, 32. Cambridge, MA: MIT Press.

Rawai, N. M., M. S. Fathi, M. Abedi, and S. Rambat. 2013. "Cloud computing for green construction management." In *Proc., 2013 Third Int. Conf. on Intelligent System Design and Engineering Applications*, 432–435. New York: IEEE.

Ross, S., G. Gordon, and D. Bagnell. 2011. "A reduction of imitation learning and structured prediction to no-regret online learning." In *Proc., Fourteenth Int. Conf. on Artificial Intelligence and Statistics*, 627–635. Maastricht, Netherlands: ML Research Press.

RSMeans. 2016. *Building construction cost data*. Norwell, MA: Construction Publishers & Consultants.

Ryu, J., M. M. Diraneyya, C. T. Haas, and E. Abdel-Rahman. 2021. "Analysis of the limits of automated rule-based ergonomic assessment in bricklaying." *J. Constr. Eng. Manage.* 147 (2): 04020163. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001978.

Schaal, S., J. Peters, J. Nakanishi, and A. Ijspeert. 2005. "Learning movement primitives." In *Proc., Robotics Research. The Eleventh Int. Symp.: With 303 Figures*, 561–572. Berlin: Springer.

Shayesteh, S., and H. Jebelli. 2021. "Investigating the impact of construction robots autonomy level on workers' cognitive load." In *Proc., Canadian Society of Civil Engineering Annual Conf.*, 255–267. Singapore: Springer.

Shi, H., R. Li, X. Bai, Y. Zhang, L. Min, D. Wang, X. Lu, Y. Yan, and Y. Lei. 2023. "A review for control theory and condition monitoring on construction robots." *J. Field Rob.* 40 (4): 934–954. https://doi.org/10.1002/rob.22156.

Sing, M. C., V. W. Tam, I. W. Fung, and H. J. Liu. 2017. "Critical analysis of construction workforce sustainability in a developed economy–case study in Hong Kong." *Proc. Inst. Civ. Eng. Eng. Sustainability* 171 (7): 342–350. https://doi.org/10.1680/jensu.17.00007.

Sokas, R. K., X. S. Dong, and C. T. Cain. 2019. "Building a sustainable construction workforce." *Int. J. Environ. Res. Public Health* 16 (21): 4202. https://doi.org/10.3390/ijerph16214202.

Sun, Y., I. Jeelani, and M. Gheisari. 2023. "Safe human-robot collaboration in construction: A conceptual perspective." *J. Saf. Res.* 86 (Sep): 39–51. https://doi.org/10.1016/j.jsr.2023.06.006.

Tolani, D., A. Goswami, and N. I. Badler. 2000. "Real-time inverse kinematics techniques for anthropomorphic limbs." *Graphical Models* 62 (5): 353–388. https://doi.org/10.1006/gmod.2000.0528.

Wan, W., Y. He, J. Liu, H. Lu, and H. Zhang. 2020. "Application of 'BIM+' architecture based on cloud technology in intelligent management of rail transit." In *Resilience and sustainable transportation systems*, 474–484. Reston, VA: ASCE.

Wang, X., C. J. Liang, C. C. Menassa, and V. R. Kamat. 2021a. "Interactive and immersive process-level digital twin for collaborative human–robot construction work." *J. Comput. Civ. Eng.* 35 (6): 04021023. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000988.

Wang, X., S. Wang, C. C. Menassa, V. R. Kamat, and W. McGee. 2023a. "Automatic high-level motion sequencing methods for enabling multi-tasking construction robots." *Autom. Constr.* 155 (Aug): 105071. https://doi.org/10.1016/j.autcon.2023.105071.

Wang, X., H. Yu, W. McGee, C. C. Menassa, and V. R. Kamat. 2023b. "Enabling BIM-driven robotic construction workflows with closed-loop digital twins." Preprint, submitted June 16, 2023. http://arxiv.org/abs/2306.09639.

Wang, Y., C. C. Beltran-Hernandez, W. Wan, and K. Harada. 2021b. "Robotic imitation of human assembly skills using hybrid trajectory and force learning." In *Proc., 2021 IEEE Int. Conf. on Robotics and Automation (ICRA)*, 11278–11284. New York: IEEE.

Wang, Z., Y. Gan, and X. Dai. 2020. "An environment state perception method based on knowledge representation in dual-arm robot assembly tasks." *Int. J. Intell. Rob. Appl.* 4 (2): 177–190. https://doi.org/10.1007/s41315-020-00128-1.

Wilder, R. 2013. "Big increase in Gulf coast projects equals big demand for skilled workers." *The Conrnerstone*. Northbrook, IL: Scott Foresman.

Wos, P., and R. Dindorf. 2023. "Develop and Implement a masonry algorithm control in a bricklaying robot." In Vol. 2949 of *Proc., AIP Conf.* Woodbury, NY: American Institute of Physics.

Wu, H., H. Li, X. Fang, and X. Luo. 2022. "A survey on teaching workplace skills to construction robots." *Expert Syst. Appl.* 205 (Nov): 117658. https://doi.org/10.1016/j.eswa.2022.117658.

Xie, F., A. Chowdhury, M. De Paolis Kaluza, L. Zhao, L. Wong, and R. Yu. 2020. "Deep imitation learning for bimanual robotic manipulation." In Vol. 33 of *Proc., Advances in Neural Information Processing Systems*, 2327–2337. Cambridge, MA: MIT Press.

Xu, G., M. Li, C. H. Chen, and Y. Wei. 2018. "Cloud asset-enabled integrated IoT platform for lean prefabricated construction." *Autom. Constr.* 93 (Sep): 123–134. https://doi.org/10.1016/j.autcon.2018.05.012.

Xu, L., C. Feng, V. R. Kamat, and C. C. Menassa. 2020. "A scene-adaptive descriptor for visual SLAM-based locating applications in built environments." *Autom. Constr.* 112 (Apr): 103067. https://doi.org/10.1016/j.autcon.2019.103067.

Yamada, T., J. Tatsuno, and H. Kobayashi. 2001. "A practical way to apply the natural human like communication to human-robot interface." In *Proc., 10th IEEE Int. Workshop on Robot and Human Interactive Communication. ROMAN 2001 (Cat. No. 01TH8591)*, 158–163). New York: IEEE.

Yang, S., P. Liu, and N. E. Pears. 2023. "Benchmarking of robot arm motion planning in cluttered environments." In *Proc., 28th Int. Conf. on Automation and Computing (ICAC2023)*. New York: IEEE.

Yang, S., X. Yu, and Y. Zhou. 2020. "LSTM and GRU neural network performance comparison study: Taking yelp review dataset as an example." In *Proc., 2020 Int. Workshop on Electronic Communication and Artificial Intelligence (IWECAI)*, 98–101. New York: IEEE.

Yu, H., V. R. Kamat, C. C. Menassa, W. McGee, Y. Guo, and H. Lee. 2023a. "Grip state recognition for enabling safe human-robot object handover in physically collaborative construction work." In *Computing in civil engineering 2023*, 787–795. Reston, VA: ASCE.

Yu, H., V. R. Kamat, C. C. Menassa, W. McGee, Y. Guo, and H. Lee. 2023b. "Mutual physical state-aware object handover in full-contact collaborative human-robot construction work." *Autom. Constr.* 150 (Apr): 104829. https://doi.org/10.1016/j.autcon.2023.104829.

Zha, Y., S. Bhambri, and L. Guan. 2021. "Contrastively learning visual attention as affordance cues from demonstrations for robotic grasping." In *Proc., 2021 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 7835–7842. New York: IEEE.

Zhang, D., Q. Li, Y. Zheng, L. Wei, D. Zhang, and Z. Zhang. 2021. "Explainable hierarchical imitation learning for robotic drink pouring." *IEEE Trans. Autom. Sci. Eng.* 19 (4): 3871–3887. https://doi.org/10.1109/TASE.2021.3138280.

Zhu, Q., T. Zhou, P. Xia, and J. Du. 2022. "Robot planning for active collision avoidance in modular construction: Pipe skids example." *J. Constr. Eng. Manage.* 148 (10): 04022114. https://doi.org/10.1061/(ASCE)CO.1943-7862.0002374.