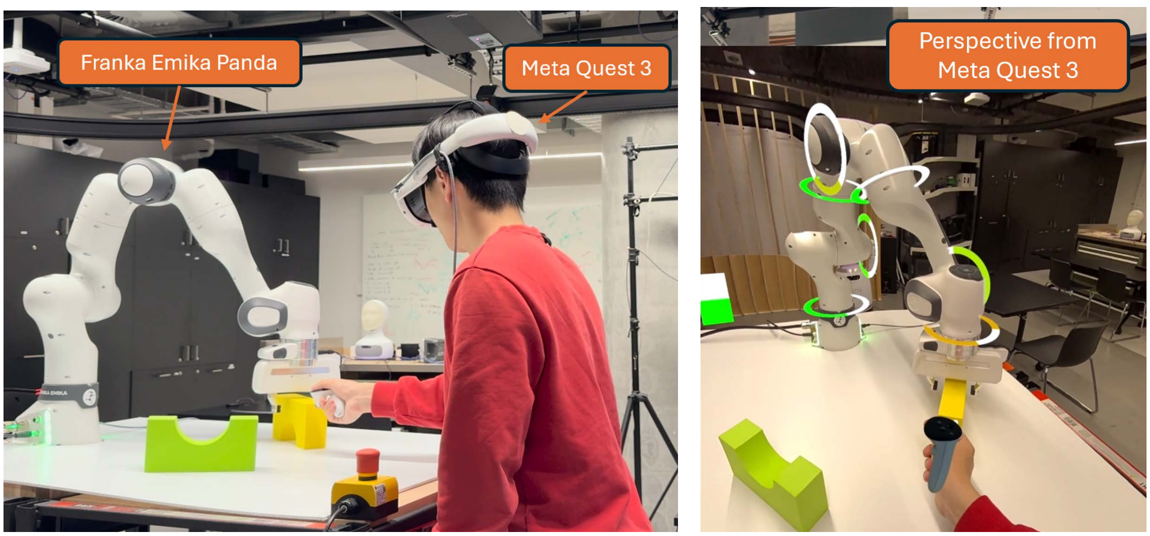
Mr.LfD: A Mixed Reality Interface for Robot Learning from Demonstration

JIAHAO CHEN, The University of Melbourne, Australia
D. ANTONY CHACON SALAS, The University of Melbourne, Australia
MUHAMMAD BILAL, The University of Melbourne, Australia
QIUSHI ZHOU, Aarhus University, Denmark
WAFA JOHAL, The University of Melbourne, Australia



(1a) User interacting with the robotic arm using Meta Quest 3.

(1b) Perspective view as seen through Meta Quest 3.

Fig. 1. A user performs an LfD task with a Franka robot using Mr.LfD through Meta Quest 3. Mr.LfD allows users to intuitively perform kinesthetic teaching or teleoperation using their hands or VR controllers, with real-time awareness of the robot's joint limits overlaid on the physical robot and its manipulability displayed beside it.

Learning from Demonstration (LfD) plays a crucial role in human-robot interaction (HRI), enabling humans to teach robots desired behaviours by demonstration. Advances in mixed-reality (MR) have introduced novel HRI techniques that overcome physical constraints and offer immersive experiences. However, MR integration with LfD remains under-explored. We present Mr.LfD, an MR interface for robot LfD that supports kinesthetic teaching and teleoperation, enhanced by 3D visualisations of the robot's status, while using

Authors' Contact Information: Jiahao Chen, The University of Melbourne, Melbourne, Australia, jiahchen4@student.unimelb.edu.au; D. Antony Chacon Salas, The University of Melbourne, Melbourne, Australia, antony.chacon@unimelb.edu.au; Muhammad Bilal, The University of Melbourne, Melbourne, Australia, muhammad.bilal1@student.unimelb.edu.au; Qiushi Zhou, Aarhus University, Aarhus, Denmark, qiushi.zhou@cs.au.dk; Wafa Johal, The University of Melbourne, Melbourne, Australia, wafa.johal@unimelb.edu.au.

Manuscript submitted to ACM

hand-tracking for real-time control. A user study measured performance through completion time and attempts, with questionnaires for feedback. Participants found our MR teleoperation interface more user-friendly than kinesthetic teaching. We found that the robot status visualisation effectively bridged the information gap, enhancing communication and maintaining user focus on the robot. We envision potential for implementing and extending Mr.LfD, and illuminate future research directions.

CCS Concepts: • **Human-centered computing** → **Interactive systems and tools**.

Additional Key Words and Phrases: Human-Robot Interaction, Learning from Demonstration, Mixed Reality, Teleoperation

1 Introduction

Recent studies in human-robot interaction (HRI) have focused on facilitating interactions between robots and users with limited programming and robotics knowledge [10, 16], because teaching robots to learn new tasks typically requires proficiency in programming, which could be time-consuming and demanding [25]. Learning from Demonstration (LfD) allows robots to acquire skills by observing or emulating human actions [10, 25], enabling domain experts to design and customise robot behaviours without needing extensive robotics knowledge.

The widespread adoption of mixed reality (MR) headsets at the consumer level has enabled novel approaches to HRI and LfD [15, 39]. MR combines virtual reality (VR) and augmented reality (AR) [33], allowing users to interact with virtual objects in the real world. Unlike traditional interactions that rely primarily on the robot's internal physical or visual feedback mechanisms, such as movements, gestures, gaze outputs, physical transformations, signal lights and small displays [22, 36, 39], MR interfaces can be designed without the constraints of the physical environment or of the robot's physical design [36], providing users with enhanced 3D visualisations [39]. For instance, Meta Quest 3 supports the DepthAPI, which generates real-time depth maps [28]. In this context, merging MR with robotics offers extensive potential for LfD through teleoperation [39], which can enhance performing demonstrations by freeing the human users from being physically present next to the robot.

Previous studies have integrated MR to improve kinesthetic teaching [15, 19, 23], a type of LfD method that relies solely on the development of the robot's hardware without the requirement of additional sensors or interfaces [25]. However, the integration of MR into robot LfD remains under explored. Prior research on robot teleoperation has not focused on integrating LfD, and teleoperation interfaces using traditional input devices like keyboards or mice may not be intuitive for novice users. Furthermore, past research on LfD has primarily focused on improving kinesthetic teaching, often within simulation environments, which still have a significant gap compared to real-world scenarios. Additionally, most past studies have used the Hololens 2. While its see-through holographic lenses offer irreplaceable advantages over camera-passthrough headsets, its chip performance, field of view (FoV), and display resolution are limited.

In this work, we present Mr.LfD, an MR interface for robot LfD that supports both kinesthetic teaching and teleoperation wth immersive visualisations. In this paper, we describe the implementation of Mr.LfD that runs natively on the Meta Quest 3 with the Franka Emika Research 3 robot. Notably, the application can operate both in a simulation environment and in the real world. Figure 1 shows the work scene of our application, with the perspective from the Quest 3 on the right. Key features of Mr.LfD include: **Real-Time Teleoperation** that offers multiple modes and is more natural for users than traditional methods like keyboards or mice, as it directly tracks the user's hands and manipulates the robotic arm in real time; **3D Visualisation** that displays the robot's internal status, such as joint positions and manipulability, to help novice users avoid crashes and facilitate demonstrations, while supporting both kinesthetic teaching and teleoperation.

Additionally, we conducted an exploratory user study to understand how Mr.LfD supports kinesthetic teaching and teleoperation for performing demonstrations and evaluate the effectiveness of the 3D visualisations. We found that teleoperation using Mr.LfD has similar performance to kinesthetic teaching, while novice users prefer the former as it is less physically demanding. Moreover, in both manipulation methods, the 3D visualisations of robot status helped users to avoid exceeding the robot's joint limits, leading to more efficient demonstration creation.

In summary, our contribution is twofold: (1) Mr.LfD, a novel interface that leverages MR to improve user experience of robot LfD; and (2) present an exploratory user study that provides insights into user experience with Mr.LfD and inspires future improvements and research directions. The documentation and source code for Mr.LfD are openly available on GitHub: https://github.com/CHRI-Lab/MrLfD_Hub.

2 Background

2.1 Mixed Reality in Robotics

Existing methods for HRI can be categorised based on the taxonomy by Phaijit et al. [22] into the following types: (1) kinesthetic interaction [26], (2) graphical user interface (GUI), (3) teleoperation, (4) Internet of Things (IoT) mediated interaction [30], (5) simulation, (6) VR, and (7) MR. MR technologies, which cover VR and AR, can be integrated to enhance the shared perception of users and robots, and can be classified based on the visualisation items location: on-robot, on-body, and on-environment [36]. On-robot involves adding information to robots through AR devices, like cartoon faces to show emotions [43] or displaying a robotic arm's motion intent [27]. On-body anchors virtual entities to users, such as a virtual monitor displaying the robot's camera perspective during teleoperation [11]. On-environment augments the surrounding scene, which includes other objects besides the robot. Examples include presenting information on large surfaces around robots [8] or enabling interactions with virtual objects, like rendering virtual replicas in manipulation tasks [7].

MR seamlessly blends virtual and physical environments, enhancing HRI through various modalities. These modalities encompass (1) tangible interaction, where users directly engage with physical objects, such as guiding robots through movements, as employed in LfD [25]; (2) touch interfaces, enabled by touch screens or virtual displays within MR setups, providing precise control over robot actions; (3) spatial interaction facilitated by pointers and controllers, the latter offering haptic feedback for optimised manipulation; (4) hand gestures enabling actions like selection, grabbing, dragging, and zooming; (5) gaze interaction, often paired with hand gestures and now prevalent in most headsets; (6) voice input facilitating command delivery; and (7) human proximity detection, empowering robots to dynamically adjust their behaviour, thereby ensuring safety, which is particularly important in the context of collaborative robots [41]. Mr.LfD incorporates MR to visualise the robot's status and enable users to teleoperate the robot using hand tracking.

2.2 Learning from Demonstration

LfD, also known as imitation learning or behavioural cloning, has attracted significant research interest over the past decade, offering a way for users without robotics expertise to program robots. It is framed as a supervised learning problem based on demonstrations provided by human teachers, and it can be categorised into three types: kinesthetic teaching, teleoperation, and observation [25].

In kinesthetic teaching, users provide demonstrations by physically interacting with the robot to teach the desired tasks [3]. However, non-expert users often provide sub-optimal demonstrations [25], which can potentially impact the robot's ability to efficiently learn and execute the tasks [2], necessitating the expertise of the demonstrators.

In teleoperation, robots are manipulated using external devices such as controllers [44]. Compared to kinesthetic teaching, teleoperation does not require users to be physically co-located with the robots, allowing for remote demonstrations. Additionally, it can be applied to more complex platforms like humanoids [44] and robotic hands [1].

Finally, in observation, robots passively observe the user's actions during the demonstration [6]. The user performs the desired task through their body movements or by wearing additional sensors for tracking, eliminating the need for expertise in robot manipulation. In our implementation of Mr.LfD, we focus on improving kinesthetic teaching and teleoperation through mixed reality.

2.3 Mixed Reality in Learning from Demonstration

The integration of MR techniques into robot LfD remains a novel and emerging field, possibly due to the historically poor performance of MR devices. Previous research has explored combining MR technologies with kinesthetic teaching, such as visualising robot constraints and acquired skills in situ [15, 16, 29]. Another effective approach for enhancing robot LfD involves implementing user interface (UI) interventions that enable teachers to receive real-time feedback from the robot regarding its learning status and make adjustments. AR interfaces have been proven to outperform traditional devices like tablets by overlaying information on the physical world and providing a semantic explanation for each action [5], offering a more immersive real-world experience. These UI interventions can be designed to minimise the time required for creating demonstrations and to enhance the overall efficiency of the teaching process.

Kinesthetic teaching with robots is generally straightforward, as it involves direct physical interaction. In contrast, teleoperation requires an additional human-machine interface to control the robot. The goal of teleoperation is to assist operators in completing complex tasks in uncertain or hazardous environments. It can also be used for demonstrations in robot LfD, allowing users to demonstrate tasks remotely without directly touching or being co-located with the robot. Various devices facilitate robotic arm teleoperation, including keyboards and mice [45], gamepads [12], mobile phones [17], 3D robotic haptic controllers [14, 20, 31], and MR headsets.

Recent advancements in MR interfaces enable integration with robot teleoperation in two primary ways: (1) serving as displays to stream live camera feeds from robots [11] and (2) tracking hand and controller movements for direct teleoperation. Commercial MR headsets have seen notable improvements in display resolution and hand tracking accuracy, enhancing both functionalities. For instance, applications like Open-TeleVision [4] demonstrate streaming capabilities by transmitting real-time stereo video feeds from humanoid robots directly to headsets like the Apple Vision Pro or Meta Quest 3, offering operators a first-person perspective during teleoperation. Conversely, our focus lies on leveraging MR headsets for precise teleoperation, akin to recent works that employed the HoloLens 2 and Vision Pro [12, 21]. Kinesthetic teaching has been shown to have the best performance in data collection compared to other manipulation interfaces implemented in previous studies, including gamepad and controller tracking, in simulation environments [12]. We aim to address this issue by deploying our teleoperation interface in a real-world implementation.

3 Mr.LfD

We introduce Mr.LfD, a mixed reality interface designed for robot LfD. This interface streamlines data collection for the Franka Research 3 robot across simulation and real-world settings through seamless integration of ROS 2 and Unity. Mr.LfD offers real-time teleoperation via hand tracking and 3D visualisation of the robot's operational status.

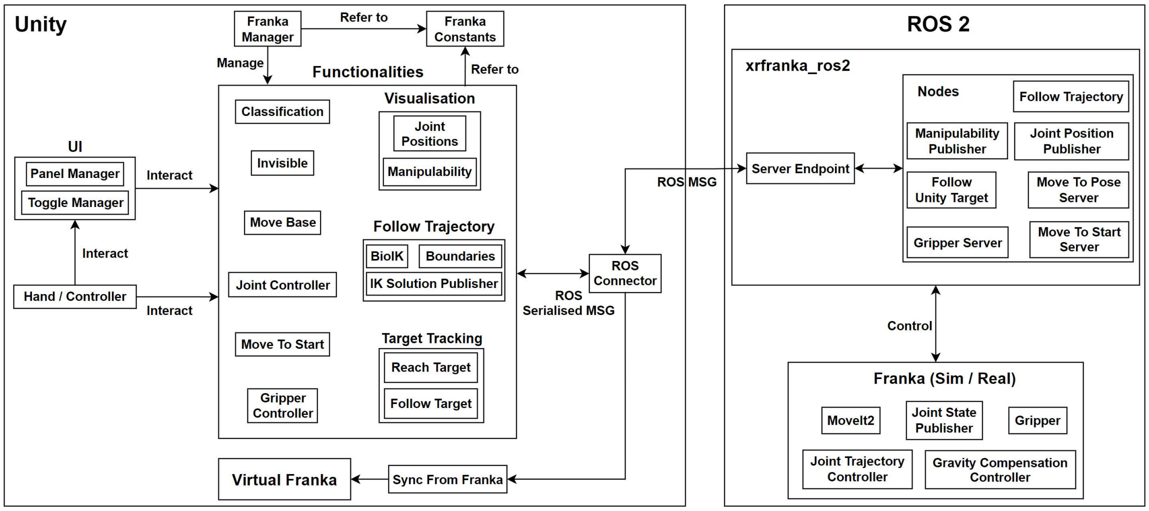


Fig. 2. System structure of Mr.LfD containing the MR part with Unity and the robot part with ROS 2 and their communication.

3.1 System Structure

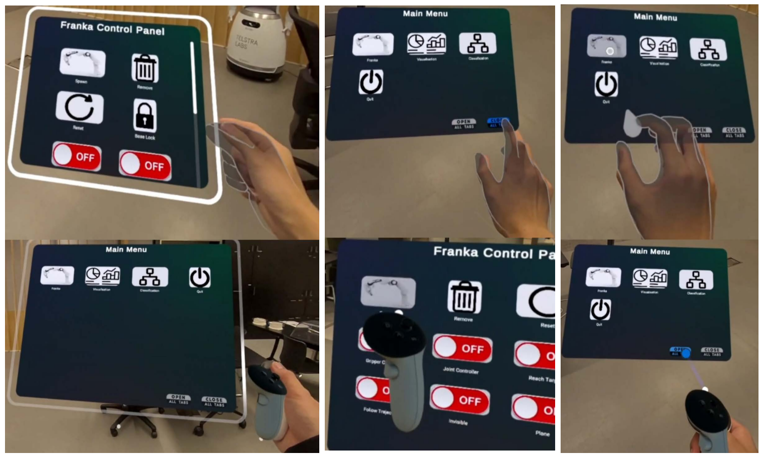
Figure 2 presents the overall structure of Mr.LfD. Users can interact with the UI and various functionalities using their hands or controllers. The communication between Unity and ROS is based on the ROS TCP Connector [37], which establishes a TCP connection, enabling the exchange of ROS messages between the virtual Franka in Unity and the real Franka controlled by ROS programs. The virtual Franka acts as a digital twin of the real Franka, facilitating bidirectional status exchange. The physical Franka's status can be synced back to the virtual Franka in Unity and vice versa, a feature essential for teleoperation and visualisation. The main components and functionalities are described as follows:

- Virtual Franka: Imported using the URDF Importer [38], which generates a 3D Franka prefab in Unity. It syncs status from the real Franka and is useful for status visualisation and teleoperation.
- Gripper Controller: Allows users to control the Franka's gripper using their hand or Quest 3's controllers.
- Joint Controller: Used to control the joints of the virtual Franka to help sync with the real one.
- Follow Trajectory: A module designed for real-time teleoperation based on an inverse kinematics (IK) solver.
- Visualisation: Visualises the robot's status in real time, including joint positions and manipulability.
- Mr.LfD_ros2: ROS 2 programs built to connect with the Unity part and control the Franka robot.

3.2 UI and Interactions

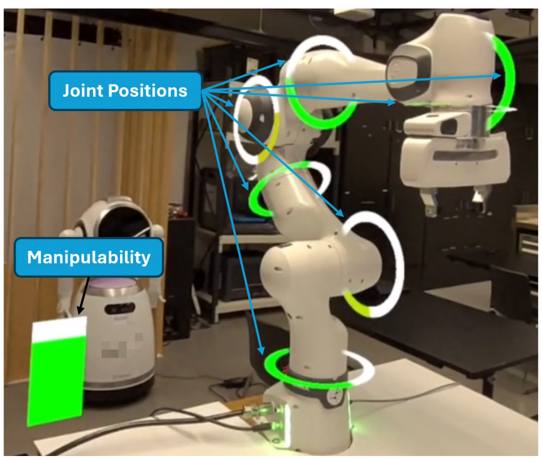
The UI panels are based on the Meta XR Interaction SDK OVR Samples [1] [18]. Figure 3 shows the Franka Control Panel, a 2D panel for navigating through menu options that affords configurations of different modes for manipulating the robot. Using their hands or controllers, users can grab the panel to modify its position, select items through manual poking or through raycasting over a distance (Figure 3).

[1] https://developer.oculus.com/downloads/package/meta-xr-interaction-sdk-ovr-samples/

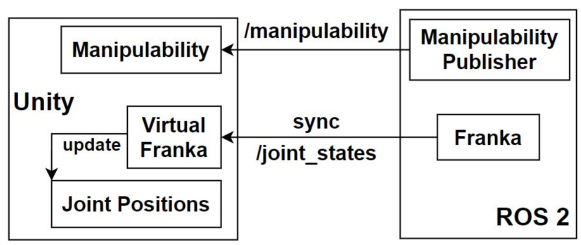


(3a) Grabbing (3b) Poking (3c) Ray Casting

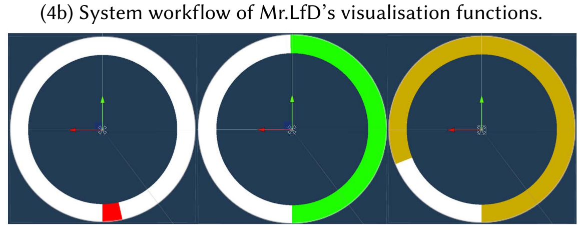
Fig. 3. UI interaction supported by the Franka Control Panel, including grabbing to adjust its position, and selection of menu items through poking or raycasting. All interactions support controller and hand gestures from both hands.



(4a) Mr.LfD visualises joint positions with reference to their limits directly on the physical robot, along with a quantitative representation of the overall manipulability by its side.



(4b) System workflow of Mr.LfD's visualisation functions.



(4c) The ring indicators dynamically change their colours according to the real-time joint positions of the robot with reference to its joint limits (from green to yellow then red when approaching limits).

Fig. 4. Overview of Mr.LfD's visualisation functions: (a) joint positions and manipulability visualisation, (b) the system workflow of these functions, and (c) colour-coded ring indicators to illustrate joint limits.

3.3 Robot Status Visualisation

For users with no background knowledge in robotics, understanding the robot's current status can be challenging, potentially leading to failed demonstrations. In Mr.LfD, we chose to visualise two types of robot status: joint positions and manipulability. Additionally, Mr.LfD features the Depth API that helps improve visualisation in 3D space. Figure 4a shows the visualisation activated on the Franka, and Figure 4b shows the workflow of this functionality.

3.3.1 Joint Positions. Franka has 7 revolute joints that are limited in their ranges of movement, exceeding which will cause Franka to stop working. Therefore, it is important to inform users of the current joint positions and alert them before approaching the joint limits. To address this, we designed ring indicators anchored at the 7 joints of the virtual Franka, updated in real-time according to its joint positions. By Syncing the virtual with the physical Franka via the ROS message /joint_states, the ring indicators accurately reflect the physical robot's joint positions. The filled colour dynamically change following the joint rotation as Figure 4c shows, to remind and alert users when their operations on the robot approach its joint limits.

3.3.2 Manipulability. Manipulability is defined as a quantitative measure of a robot's ability to change the position and orientation of its end effector [42]. Franka's end effector is defined as the centre of the two fingers of the gripper. A bar indicator is implemented to represent the manipulability measure in real time. Manipulability is a crucial motion feature used to quantify the quality of demonstrations for developing an efficient LfD model to execute desired tasks [2].

Mathematically, manipulability is represented as a scalar value derived from the Jacobian matrix of the robotic arm, indicating how easily and efficiently Franka can perform various tasks. The Jacobian matrix $\mathcal{J}$ is frequently used in robotics and control theory to represent the kinematic relationships and dynamics of robots. It is defined as:

$$J = \frac{\partial x}{\partial q} \tag{1}$$

For a robot with n joints, x represents the end effector's position and orientation in 3D space, which is a 6-dimensional vector. And q represents the joint angles and displacements, which has n dimensions. The Jacobian matrix captures the partial derivatives of the end effector's position and orientation with regard to the joint variables, resulting in a $6 \times n$ matrix. Manipulability w is measured as:

$$w = \sqrt{\det(JJ^T)} \tag{2}$$

In our implementation, it is computed using the SVD (singular value decomposition) of the Jacobian matrix and is publishd as the ROS message /manipulability:

$$J = U\Sigma V^T \tag{3}$$

Σ is a diagonal matrix of singular values $\sigma_1, \sigma_2, \cdots, \sigma_m$, and manipulability w can also be express as their product:

$$w = \sigma_1 \sigma_2 \cdots \sigma_m \tag{4}$$

3.3.3 Depth API. Mr.LfD employs the DepthAPI [28], a new feature supported exclusively on Quest 3 compared to its predecessor. It leverages the depth camera to estimate the depth of objects in the environment in real time. Figure 5 shows the comparison of different occlusion effects when hands are placed in front of virtual objects. Mr.LfD implements DepthAPI through custom shaders to correctly render the ring indicators around the physical robot and to enable intuitive hand interaction with the control panel, providing an immersive interaction experience for LfD tasks.

3.4 Trajectory Following

Our teleoperation interface offers a natural way for users to manipulate the end effector, allowing it to follow the target's trajectory in real-time. Users can grab the target and move their hands to continuously move the gripper, which rotates according to the user's hand pose. To achieve this, we used an IK solver [40] to compute the desired joint positions, placing the end effector at the required position and orientation in real-time.

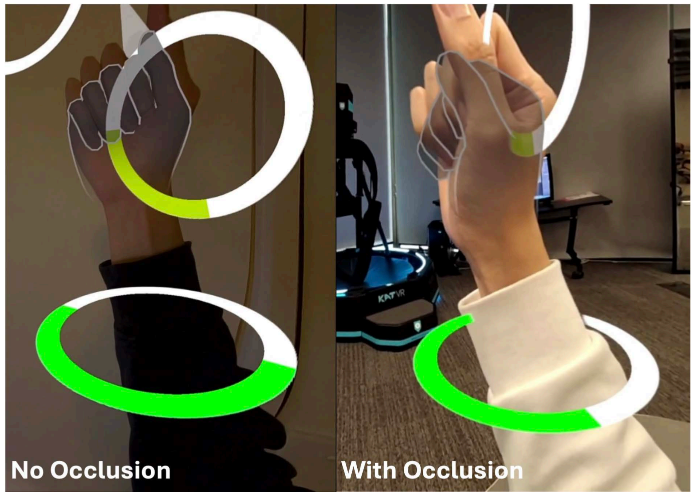


Fig. 5. Demonstration of the DepthAPI for Mr.LfD visualisation features, including ring indicators and hand interaction.

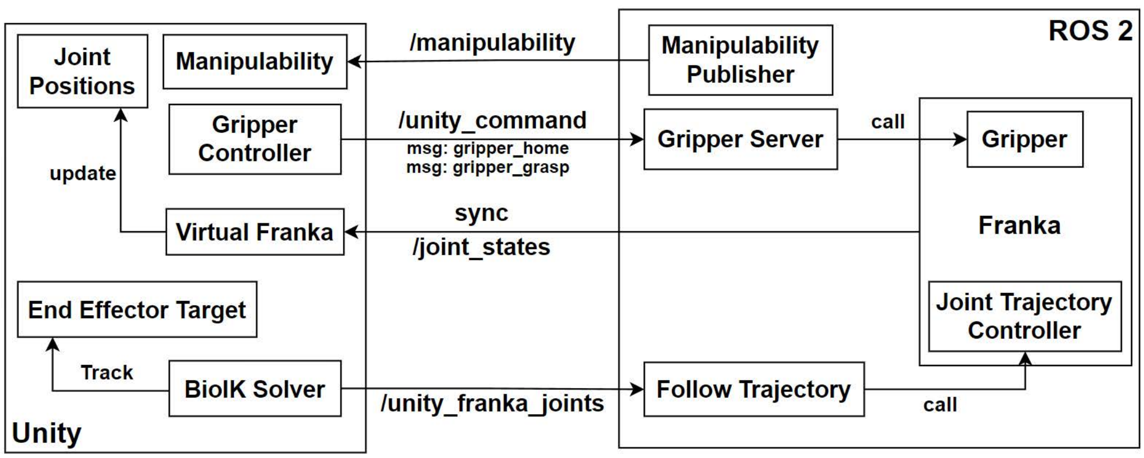


Fig. 6. Workflow of trajectory following

Figure 6 presents the workflow of trajectory following. For the IK solver, we have implemented the Hybrid Genetic Swarm Algorithm (HGSA) [35], also known as BioIK [34]. This approach integrates Genetic Algorithms (GA) and Particle Swarm Optimisation (PSO). It supports seamless integration with Unity objects and efficiently manages the kinematic chain of the Franka Research 3 robot, completing computations in under 200 ns.

The IK solutions can lead to significant differences in joint positions within a short time, which is acceptable in simulation since joint velocity has no limit. However, controlling a real robot requires careful handling of joint movements, as significant displacement between consecutive joint positions in a short time can lead to jerky movement. Therefore, we implemented a trajectory control strategy. We compute the difference between the desired joint positions and the current ones. If the absolute difference among all joints is below a threshold of 0.2 radians, the desired positions are directly sent to the joint trajectory controller in ROS2 as the goal positions; otherwise, a cosine-based trajectory generation is applied. This method generates a smooth trajectory to the goal position, minimising jerk (the rate of change of acceleration). Furthermore, we employed the joint trajectory controller with PD control [32], setting the proportional gain to 5 and a derivative gain to 50 for all joints.

Once the IK solution is computed on the Unity side, it is published back to Franka via the /unity_franka_joints ROS message to move the physical robot. Additionally, the ROS message /unity_command is used to control Franka's gripper.

4 User Evaluation

We conducted an exploratory user study to understand the user experience of LfD tasks enabled by Mr.LfD. Using the implementation of Mr.LfD described above, we instructed participants to record robot demonstrations using different features of Mr.LfD. We collected and analysed their performance data, subjective ratings of the experience, and conducted semi-structured interviews to gain insights into the usability of Mr.LfD and opportunities for future improvements. Before the user study began, a pilot study was completed by a researcher with sufficient background knowledge in robotics and Franka to ensure that the user study met safety requirements for participants with limited technical background.

4.1 Task Design

We test our system on three classic manipulation tasks: pick-and-place, object stacking, and object insertion. In order to evaluate both teleoperation and kinesthetic demonstration methods, participants are asked to perform each task twice, with robot status visualisation enabled. The teleoperation interface used is trajectory following with controller tracking (see Fig. 2). For each demonstration method and each task, participants are asked to perform three successful demonstrations, resulting in a total of 18 demonstrations per participant for the three tasks (3 tasks × 2 methods × 3 demonstrations). The number of attempts and the completion time for each successful demonstration are recorded to measure participants' performance.

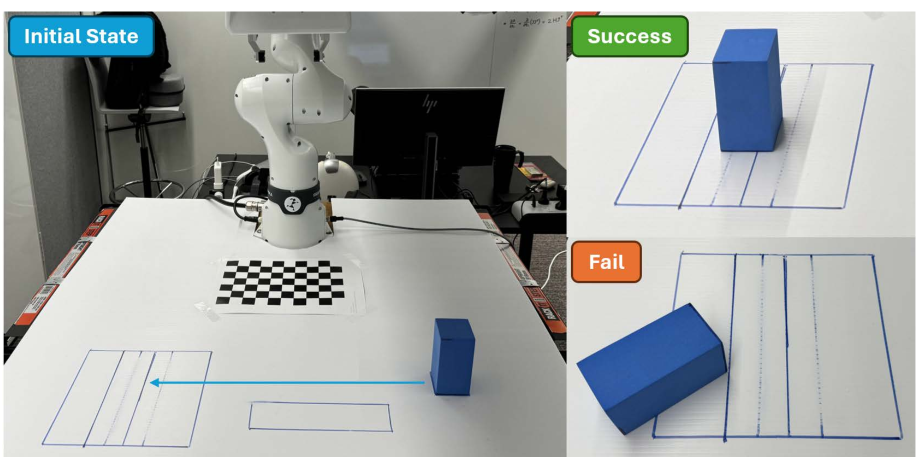


Fig. 7. Task 1: pick-and-place; using one of the demonstration method, the participant needs to operate the robot for it to grasp the blue block and place it into the goal area.

4.1.1 Pick and place. Figure 7 shows the initial and goal states of Task 1 (pick-and-place). The blue block needs to be picked up and placed into the designated square area. Placing the block anywhere within the area is considered a valid demonstration. If the block is placed on the line, it is out of bounds and is regarded as a failure.

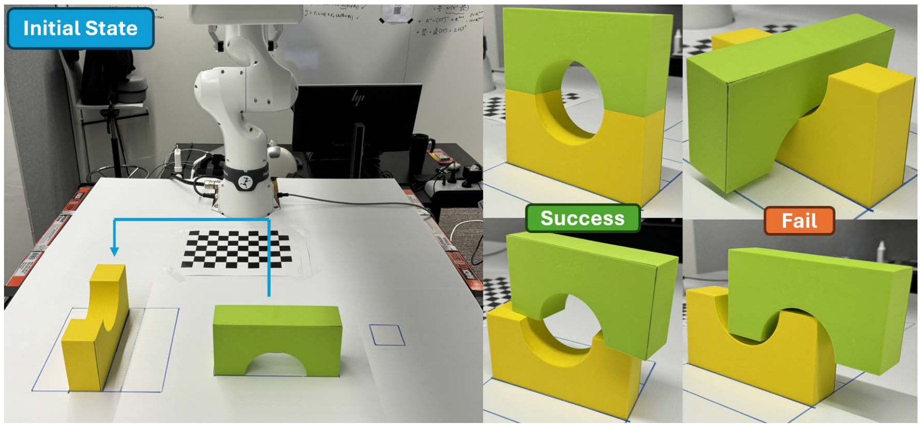


Fig. 8. Task 2: object stacking; using one of the demonstration method, the participant needs to operate the robot for it to grasp the green block and place it on top of the yellow block. The total height of the stack should be the sum of the height of the two blocks.

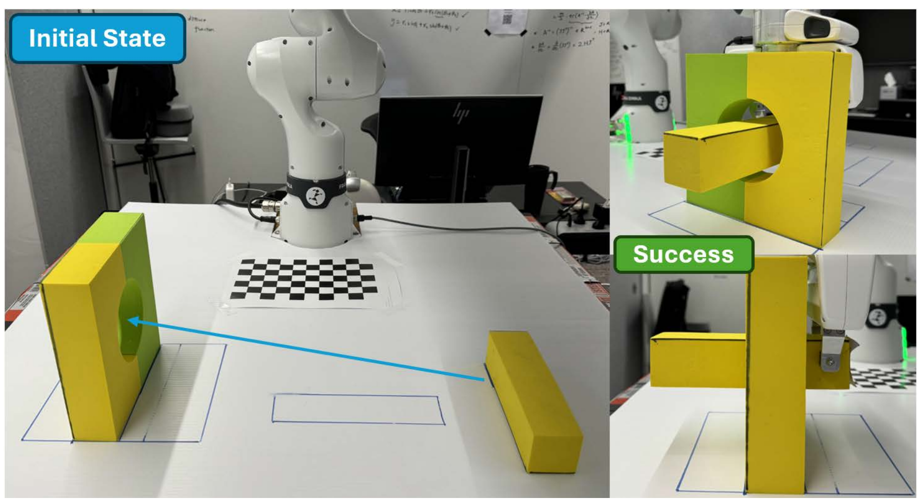


Fig. 9. Task 3: object insertion; using one of the demonstration method, the participant needs to operate the robot for it to grasp the yellow cuboid and insert it into the cavity made by the two other blocks.

4.1.2 Object stacking. Figure 8 shows the initial and goal states of Task 2 (object stacking). The green block needs to be picked up and stacked on the yellow block. The task is successful as long as the green block does not fall. However, it must be ensured that the stacking is based on two legs, creating a hollow hole in the centre of the two blocks. Other stacking approaches are not permitted and will also be regarded as failures. Additionally, if the position of the yellow block is moved during manipulation, it will be considered a failure. Compared to Task 1, which only involves translation, Task 2 requires users to also rotate the block for precise placement.

4.1.3 Object insertion. Figure 9 shows the initial and goal states of Task 3 (object insertion). The yellow cuboid block needs to be picked up and inserted into the hole in the centre of the other block. The success condition is that over half of the block has been inserted into the hole without touching the hole's edge. For success, the block should be grasped at the edge location to ensure enough length for insertion. Compared to the previous tasks, this task not only requires rotation but also much more precise manipulation to ensure insertion without collision.

For each task, participants first perform three demonstrations using the kinesthetic teaching method, followed by three demonstrations using teleoperation method. Before starting Task 1, participants are given an introduction to the Quest 3 and Franka and provided with time to practice manipulating Franka using both kinesthetic and teleoperation teaching methods through Mr.LfD. Once they feel ready, they can begin Task 1.

4.2 Questionnaire

Before the study starts, participants are required to read and sign a consent form and complete a demographic form. The demographic form collects participants' age, gender, and background information. Additionally, participants are asked to respond to four questions related to the study, using a 10-point scale, where 1 represents the lowest frequency and 10 represents the highest frequency. These questions aim to gather information on participants' backgrounds in terms of usage and knowledge in video games, AR/VR headsets, robotic arms, and activities requiring upper limb coordination.

After all demonstrations for a task are completed, participants are asked to fill a form to assess their experience. We first asks participants to rate the task's difficulty for both kinesthetic teaching and teleoperation on a scale of 10. Participants' cognitive demand during manipulation is collected using NASA-TLX on a 20-point scale for each dimension [9].

Finally, participants were asked to rate their agreement on a 10-point scale to assess their satisfaction with the responsiveness of the system when teleoperating it and to measure the effect of robot status visualisations in both manipulation methods. At the end of the study, participants were invited to provide more comments.

4.3 Participants

We recruited 10 right-handed participants ($3F$, $7M$), with ages ranging from 21 to 30 years (mean = 25.2, SD = 2.62). Most participants were novices in AR and VR technologies and robotics.

4.4 Study Procedure

The user study was conducted in a User Experience Lab. The experimenter was seated at a desk behind the setup to manage the ROS programs and monitor the study. A separate laptop streamed the participants' first-person view from the Quest 3, allowing the experimenter to give instructions effectively. Additionally, an emergency button was placed at the experimenter's desk to handle any unexpected errors that could pose risks to participants and the robotic arm.

During the study, participants were asked to manipulate Franka and perform three tasks (pick-and-place, object stacking and insertion). After completing each task, participants fill out the questionnaire to record their user experience. The study took approximately 90 minutes for each participant.

5 Results

5.1 Quantitative Results

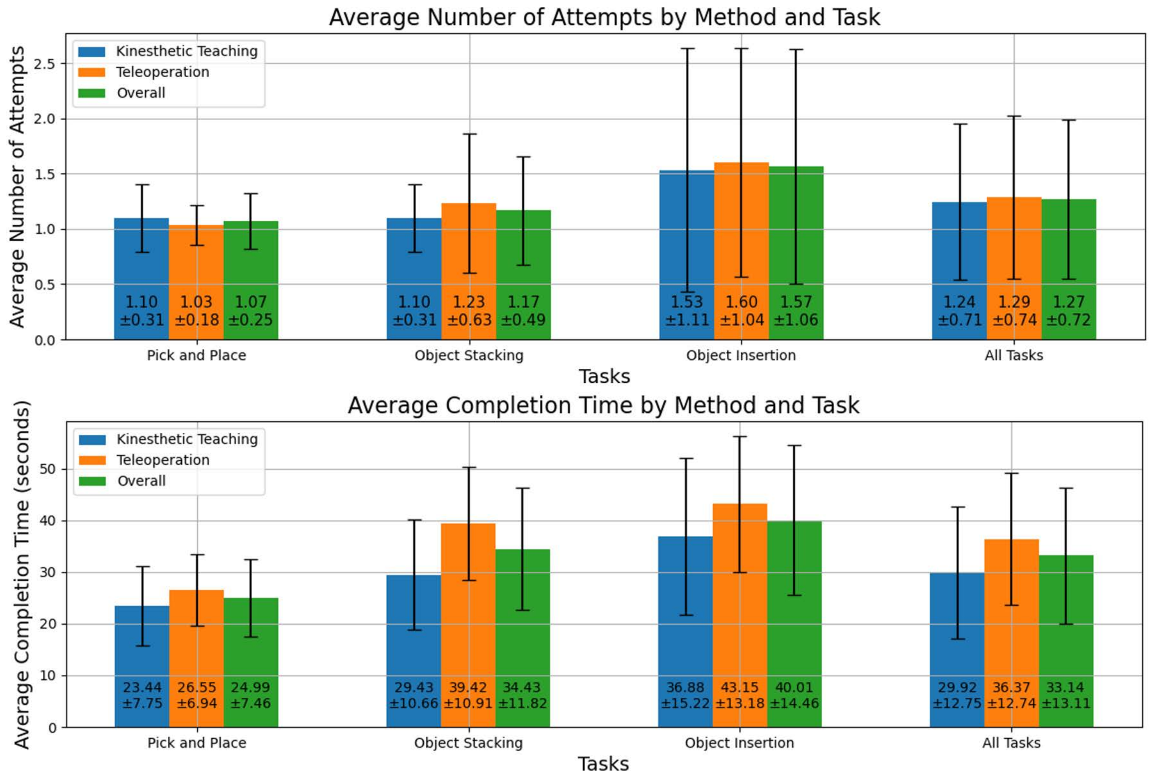


Fig. 10. Average number of attempts and completion time, categorised by manipulation method and task

5.1.1 *Performance in Kinesthetic Teaching vs. Teleoperation.*

A total of 180 successful demonstrations were collected (60 per task). To evaluate participants' performance, the number of attempts required for each successful demonstration and the completion time for each successful attempt were recorded. Figure 10 presents the average number of attempts and completion time, categorised by manipulation method and task ($N = 30$ for each method and each task). Descriptively looking at the data, we did not see a difference between the manipulation method. We do observe that, as per our design, the object insertion task seemed more difficult than the two other tasks and that, as expected, the pick-and-place task seemed to be the easiest.

5.1.2 *Usability of Kinesthetic Teaching vs. Teleoperation.*

Figure 11 presents the mean NASA-TLX subscores for all tasks manipulated using the two methods. The task load index is computed as the average of the six dimensions of NASA-TLX.

Descriptively analysing the data, there is no overall observable difference between the two methods. Kinesthetic teaching resulted in lower mental demand, effort, and frustration levels, while teleoperation reduced physical demand and made participants feel more successful in their performance. These results suggest that our teleoperation interface offers usability at least at the same level as kinesthetic teaching for performing these three tasks.

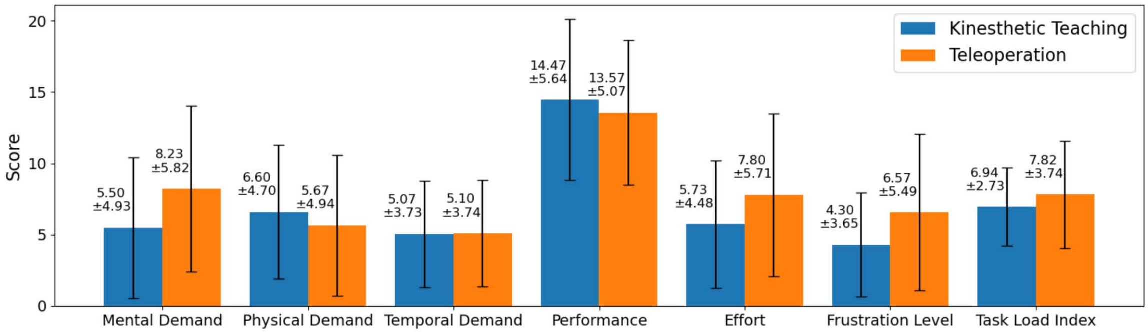


Fig. 11. Mean and standard deviation of NASA-TLX scores for all tasks performed using either kinesthetic teaching or teleoperation

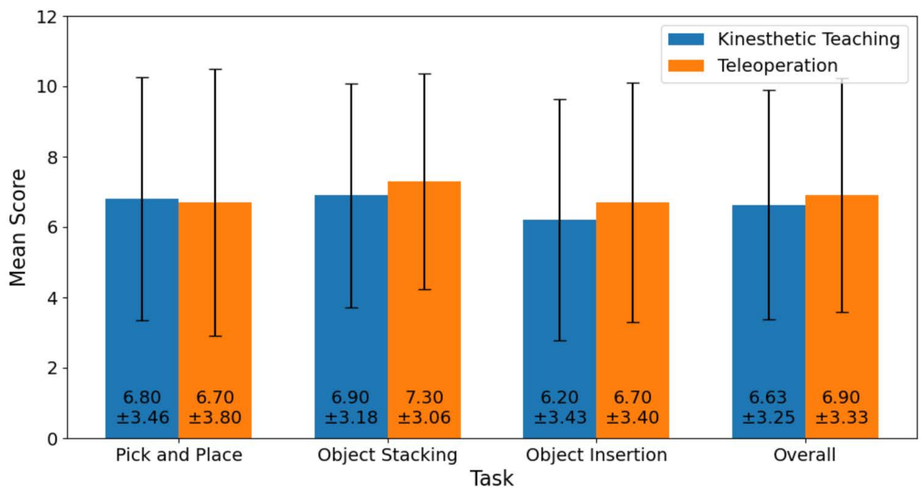


Fig. 12. Mean score on visualisation helpfulness per task

5.1.3 Usage of Robot Status Visualisation. To assess whether robot status visualisation aids participants in performing demonstrations, Figure 12 presents the mean score on visualisation helpfulness per task, as collected from the questionnaire. Again here, our study does not show a noticeable difference between the two manipulation methods.

5.2 Qualitative Results

5.2.1 Kinesthetic Teaching vs. Teleoperation. Of all the participants, two (P4 and P6) showed a preference for kinesthetic teaching, six (P1,P2,P5,P7,P8,P10) preferred teleoperation, and two (P3 and P9) held neutral opinions. Kinesthetic teaching was initially expected to be a simpler method for new users to become familiar with the system, but some participants noted that it requires higher physical exertion (P1, P2, P7, P9). Based on our study setup, the robot manipulation platform is high and wide, making it particularly physically demanding for participants of shorter stature to reach the gripper (P2, P7). Additionally, participants need to apply more force to manipulate the gripper in some configurations, which can cause anxiety due to the fear of damaging the robotic arm (P3). It is worth noting that one of the two participants who preferred kinesthetic teaching had experience with robotic arms and found this method more natural to use (P6). However, for the other participants with no prior experience in robotics, our teleoperation interface was generally preferred for these tasks, especially once they had become accustomed to its controls.

Furthermore, participants may have different preferences depending on the task. Some participants found teleoperation easier (P2, P7, P9), particularly for object insertion, which is the most challenging task requiring high-precision control. Observations revealed that many participants struggled to maintain the gripper pose when inserting objects using kinesthetic teaching, and applying too much force could easily lead to collisions, necessitating precise force control with the hands. In contrast, teleoperation only requires participants to keep their hand stable without needing to control the force applied to the gripper, as this is managed automatically by ROS programs.

5.2.2 Visualisation. While some participants utilised the visualisation to assist with manipulation, others did not use it. One participant (P5) expressed fear when the rings turned red, indicating the joints were approaching their limits. Moreover, one participant (P10) suggested adding arrows on the rings to more clearly indicate the joints' rotation direction, which is considered a beneficial feature for future implementation.

We noted that all comments from participants on the visualisation were about joint positions and joint limits. We did not get feedback on manipulability, which suggests a need for further research on the impact of visualising manipulability through different approaches and in various tasks.

6 Discussion

6.1 Performance and expertise

Despite the task order progressing from easy to difficult, allowing participants time to acclimate to the manipulation methods, they still required more attempts and longer completion times as task difficulty increased for both kinesthetic teaching and teleoperation. Performance varied significantly among participants, with those having prior experience with robotic arms demonstrating significantly better performance in both the number of attempts and completion time. Additionally, there is a noticeable negative correlation between AR/VR app usage and the number of attempts, particularly for teleoperation. This is understandable, as participants with prior AR/VR experience may quickly familiarise themselves with our application and the use of Quest 3 controllers. However, their advantage does not extend to kinesthetic teaching, which requires direct touch.

All participants successfully performed all demonstrations, with an overall average of 1.27 attempts and 33.14 seconds for completion time per round. Considering that most participants had limited prior experience with AR/VR and robotics, we demonstrate the potential of using MR headsets with camera-passthrough technologies for data collection in robot LfD. Given that commercial camera-passthrough headsets are continuously being updated, the future integration of such headsets with HRI is promising.

6.2 Kinesthetic Teaching vs. Teleoperation

The results suggest that participants have a similar success rate when performing demonstrations using either method. Observations indicate a significant increase in completion time from the pick-and-place task to the object-stacking task in teleoperation, while the time increase from object stacking to object insertion is limited. This is primarily because, in the second task, participants need to perform a 90-degree rotation to succeed, causing many to struggle with this operation using the controller. By the third task, many participants had adapted to wrist rotation, resulting in a smaller time increase than in kinesthetic teaching, despite the increased task difficulty.

The learning curves for these two methods also differ. One participant (P3) remarked, "I feel like I just started to understand the working principles for the robot (teleoperation) when the tasks are finished." Given that the overall average time difference between the two methods is about 6 seconds, which is acceptable for performing a demonstration, we can conclude that the performance of kinesthetic teaching and teleoperation is comparable for these three tasks.

It is interesting to note that most participants (7 of 10) prefer teleoperation over kinesthetic teaching from interview, despite the latter being generally favoured by roboticists. A key advantage of kinesthetic teaching is that users can control not only the end effector but also all other joints. While managing all joints simultaneously is challenging, users can choose to move different joints at different times. Currently, our teleoperation interface only supports users in controlling one joint, the end effector, while the movement of other joints is computed by the IK solver. The advantages and disadvantages of kinesthetic teaching and our teleoperation interface can be summarised as follows.

One of the primary advantages of kinesthetic teaching is that it allows users to control all the joints of the robot, generating smooth trajectories, which is particularly beneficial for robotic experts when performing demonstrations. However, there are notable drawbacks. Kinesthetic teaching may not be user-friendly for novices, who might struggle if the robot is in a state of low manipulability and may not know how to correct it. Additionally, this method requires more physical exertion from the user.

In contrast, teleoperation offers several advantages. It requires less physical exertion, which can be a significant benefit. This method can be easier and safer for novice users. Furthermore, users do not need to be concerned about low manipulability states since the IK solver can manage these situations. Despite these benefits, there are limitations. Currently, teleoperation only supports controlling the end effector, and other joints are not available for direct control.

We conclude that our teleoperation interface is more user-friendly for novice users in the tasks within our study. However, this conclusion may not extend to other tasks. Compared to daily life activities, our tasks are simpler. More advanced tasks, such as scooping or cooking, and handling soft objects like clothes, are significantly more challenging and could require controlling multiple joints simultaneously for a good demonstration.

It is also controversial whether LfD should allow auxiliary tools to manipulate the robot, as LfD aims to learn entirely from users. Using an IK solver to assist with joint movement might be seen as somewhat conflicting. However, for some tasks, only the position of the end effector needs to be learned. In our teleoperation interface, users can fully control the translation and rotation of the gripper. This approach can still be considered a type of LfD, as it allows the robotic arm to learn the desired end effector positions for these tasks. Additionally, the IK solver helps users avoid concerns about low manipulability. The ultimate goal of LfD is to enable novice users to teach robots. If external tools can aid them in performing demonstrations more effectively, then their use is justified.

6.3 Robot status visualisation

The results suggest that visualisation aids participants in task completion to some degree. However, there is considerable variation among participants. Some did not rely on the visualisations, as they quickly became familiar with the robotic arm. Conversely, others depended more on the visualisation of joint positions to ensure that joint limits were not exceeded. Additionally, some participants found the visualisation of joint positions more beneficial for the second and third tasks due to the rotational requirements involved.

Interestingly, almost no participants relied on the visualisation of manipulability. Several reasons may explain this observation. First, the current visualisation for manipulability is rather simple. Second, participants without a background in robotics may find manipulability challenging to understand. Third, participants may focus on completing the task and not pay much attention to manipulability. Finally, the task properties may make it difficult to place the robotic arm into a singular configuration. Furthermore, for teleoperation, the use of an IK solver means participants do not need to concern themselves with maintaining manipulability.

6.4 Limitations

Our user study setup could potentially limit the performance of some users in kinesthetic teaching. The manipulation platform is too large and high for users with shorter stature. It would be better to mount the Franka robot on an adjustable desk in the future to accommodate different users.

Because the user study takes a relatively long time (90 minutes), some participants could become impatient over time, which might affect their performance. For participants who easily get motion sickness, their performance on later tasks could be significantly impacted. This could be alleviated by recruiting more participants, with each performing fewer demonstrations. However, since there is a learning curve for manipulation methods, achieving a balance between becoming familiar with the system and avoiding impatience or motion sickness is important. Moreover, given the limited sample size of our study and the use of only Quest 3 and Franka Emika Research 3, the conclusions drawn can only provide insights for the integration of MR and robot LfD. It is necessary to conduct more studies with larger populations and other equipment to obtain more comprehensive evaluations.

7 Future Work

7.1 Improvements on Mr.LfD

Currently, the visualisation of joint positions has demonstrated more apparent benefits for users compared to manipulability. It is necessary to create a more vivid visualisation of manipulability, such as displaying the manipulability ellipsoid. Ideally, the visualisation should assist novice users in adjusting joint configurations when the robotic arm is in a state of low manipulability, perhaps by displaying an animation that shows how to correct the current status. As a result, the quality of demonstrations will improve, as it significantly depends on the manipulability feature [2].

It is also worth noting that teleoperation with an MR headset may lack haptic feedback, particularly when using hand tracking. This contrasts with devices like 3D robotic haptic controllers, which can provide operators with an experience as if they are directly interacting with the robot environment [24] and can improve coordination between humans and robots [14]. Possible solutions include creating a mismatch between the visual input (robot movement) and the user's proprioceptive feedback (input movement) [24] or incorporating wearable force sensors [13].

For our teleoperation, a key improvement is implementing force feedback to inform users if an object is grasped or to indicate the object's weight, providing a more immersive experience. One solution could involve leveraging the haptic motor on Quest 3 controllers to generate different haptic feedback for various situations, such as when an object is being grabbed or in response to changes in the robot's torque. Another improvement is to implement a stabilisation strategy to assist users, particularly those who struggle to keep their hands steady in the air, in performing demonstrations. However, this could potentially reduce the system's responsiveness, so allowing users to adjust the stability parameters based on their preferences would be beneficial. Additionally, while the IK solver is currently implemented on the Unity side and the teleoperation is responsive, it would be worthwhile to test implementing the IK solver on the ROS side. This could reduce latency and potentially enhance the user experience.

7.2 Research Directions

One direction is to train the data collected from our application for robot LfD. Investigating the model performance trained on data collected via kinesthetic teaching and teleoperation is essential. Additionally, for LfD, mounting a camera on the gripper to provide the robot with visual information would be ideal. Implementing an interface in our application to stream the gripper camera view, allowing users to have a first-person perspective from the robot, would be useful. Including more advanced tasks could also provide a more comprehensive evaluation of the two manipulation methods. Currently, demonstration recording occurs on the ROS side. In the future, integrating this functionality into Unity would allow users to directly manage demonstrations in Quest 3, eliminating the need for another person to operate the ROS programs.

Another idea is to enable multiple joint control in teleoperation, addressing its most significant shortcoming. Technically, this requires a more complex implementation of the IK solver to handle joints other than the end effector and may introduce more motion issues. Furthermore, Mr.LfD currently supports controlling only one Franka robot. However, enabling control of two robotic arms simultaneously using both hands could facilitate tasks requiring robot collaboration, such as transmitting objects. This is technically feasible for our system, but careful consideration must be given to preventing collisions between the two robots.

8 Conclusion

This paper introduced Mr.LfD, a novel MR interface for robot LfD that integrates kinesthetic teaching and teleoperation with immersive 3D visualisations. Our implementation on the Meta Quest 3, combined with the Franka Emika Research 3 robot, demonstrates the potential of MR to enhance user experience and improve the efficiency of robot LfD. The exploratory user study highlighted the advantages of Mr.LfD, particularly in teleoperation, which users found more intuitive and less physically demanding compared to traditional methods. The 3D visualisations provided critical feedback, helping users avoid operational errors and improving the overall demonstration process.

Our findings underscore the potential of MR in simplifying robot LfD, making it accessible to users with limited robotics expertise. The open-source release of Mr.LfD and the study data aims to foster further research and development in this field, ultimately contributing to more user-friendly and efficient human-robot interaction systems.

Acknowledgments

This work was partially supported by the Australian Research Council (Grant No. DE210100858).

References

[1] Jacopo Aleotti and Stefano Caselli. 2011. Part-based robot grasp planning from human demonstration. In *2011 IEEE International Conference on Robotics and Automation*. IEEE, 4554–4560.

[2] Muhammad Bilal, Nir Lipovetzky, Denny Oetomo, and Wafa Wafa Johalaand Johal. 2024. Beyond Success: Quantifying Demonstration Quality in Learning from Demonstration. In *2024 IEEE/RSJ International conference on Intelligent Robots and Systems (IROS)*. IEEE.

[3] Sylvain Calinon, Florent Guenter, and Aude Billard. 2007. On learning, representing, and generalizing a task in a humanoid robot. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 37, 2 (2007), 286–298.

[4] Xuxin Cheng, Jialong Li, Shiqi Yang, Ge Yang, and Xiaolong Wang. 2024. Open-TeleVision: open-source tele-operation with vision.

[5] Maximilian Diehl, Alexander Plopski, Hirokazu Kato, and Karinne Ramirez-Amaro. 2020. Augmented reality interface to verify robot learning. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 378–383.

[6] Rüdiger Dillmann. 2004. Teaching and learning of robot tasks via observation of human performance. *Robotics and Autonomous Systems* 47, 2-3 (2004), 109–116.

[7] Jared A Frank, Matthew Moorhead, and Vikram Kapila. 2016. Realizing mixed-reality environments with tablets for intuitive human-robot collaboration for object manipulation tasks. In *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*.

IEEE, 302–307.

[8] Cheng Guo, James Everett Young, and Ehud Sharlin. 2009. Touch and toys: new techniques for interaction with a remote group of robots. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 491–500.

[9] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.

[10] Bradley Hayes and Brian Scassellati. 2013. Challenges in shared-environment human-robot collaboration. *learning* 8, 9 (2013).

[11] Hooman Hedayati, Michael Walker, and Daniel Szafir. 2018. Improving collocated robot teleoperation with augmented reality. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. 78–86.

[12] Xinkai Jiang, Paul Mattes, Xiaogang Jia, Nicolas Schreiber, Gerhard Neumann, and Rudolf Lioutikov. 2024. A Comprehensive User Study on Augmented Reality-Based Data Collection Interfaces for Robot Learning. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*. 333–342.

[13] Hyung-il Kim, Boram Yoon, Seo Young Oh, and Woontack Woo. 2023. Visualizing Hand Force with Wearable Muscle Sensing for Enhanced Mixed Reality Remote Collaboration. *IEEE Transactions on Visualization and Computer Graphics* (2023).

[14] Yiming Liu, Raz Leib, and David W Franklin. 2023. Follow the force: Haptic communication enhances coordination in physical Human-Robot interaction when humans are followers. *IEEE Robotics and Automation Letters* (2023).

[15] Matthew B Luebbers, Connor Brooks, Minjae John Kim, Daniel Szafir, and Bradley Hayes. 2019. Augmented reality interface for constrained learning from demonstration. In *Proceedings of the 2nd International Workshop on Virtual, Augmented and Mixed Reality for HRI (VAM-HRI)*.

[16] Matthew B Luebbers, Connor Brooks, Carl L Mueller, Daniel Szafir, and Bradley Hayes. 2021. Arc-lfd: Using augmented reality for interactive long-term robot skill maintenance via constrained learning from demonstration. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 3794–3800.

[17] Ajay Mandlekar, Yuke Zhu, Animesh Garg, Jonathan Booher, Max Spero, Albert Tung, Julian Gao, John Emmons, Anchit Gupta, Emre Orbay, et al. 2018. Roboturk: A crowdsourcing platform for robotic skill learning through imitation. In *Conference on Robot Learning*. PMLR, 879–893.

[18] Meta. 2024. Meta XR Interaction SDK OVR Samples. Accessed: May 26, 2024.

[19] Carl Mueller, Jeff Venicx, and Bradley Hayes. 2018. Robust robot learning from demonstration and skill repair using conceptual constraints. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 6029–6036.

[20] Jiahe Pan, Jonathan Eden, Denny Oetomo, and Wafa Johal. 2024. Effects of Shared Control on Cognitive Load and Trust in Teleoperated Trajectory Tracking. *IEEE Robotics and Automation Letters* (2024).

[21] Younghyo Park and Pulkit Agrawal. 2024. Using Apple Vision Pro to Train and Control Robots.

[22] Ornnalin Phaijit, Mohammad Obaid, Claude Sammut, and Wafa Johal. 2022. A Taxonomy of Functional Augmented Reality for Human-Robot Interaction. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 294–303.

[23] Ornnalin Phaijit, Claude Sammut, and Wafa Johal. 2023. User Interface Interventions for Improving Robot Learning from Demonstration. In *Proceedings of the 11th International Conference on Human-Agent Interaction*. 152–161.

[24] Pragathi Praveena, Daniel Rakita, Bilge Mutlu, and Michael Gleicher. 2020. Supporting perception of weight through motion-induced sensory conflicts in robot teleoperation. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-robot Interaction*. 509–517.

[25] Harish Ravichandar, Athanasios S Polydoros, Sonia Chernova, and Aude Billard. 2020. Recent advances in robot learning from demonstration. *Annual review of control, robotics, and autonomous systems* 3 (2020), 297–330.

[26] Kyle B Reed, Michael Peshkin, Mitra J Hartmann, J Edward Colgate, and James Patton. 2005. Kinesthetic interaction. In *9th International Conference on Rehabilitation Robotics, 2005. ICORR 2005*. IEEE, 569–574.

[27] Eric Rosen, David Whitney, Elizabeth Phillips, Gary Chien, James Tompkin, George Konidaris, and Stefanie Tellex. 2020. Communicating robot arm motion intent through mixed reality head-mounted displays. In *Robotics Research: The 18th International Symposium ISRR*. Springer, 301–316.

[28] Oculus Samples. 2024. Depth API. Accessed: May 25, 2024.

[29] Sarah Schömbs, Jorge Goncalves, and Wafa Johal. 2024. Exploring Data Agency and Autonomous Agents as Embodied Data Visualizations. *arXiv preprint arXiv:2402.04598* (2024).

[30] Pieter Simoens, Mauro Dragone, and Alessandro Saffiotti. 2018. The Internet of Robotic Things: A review of concept, added value and applications. *International Journal of Advanced Robotic Systems* 15, 1 (2018), 1729881418759424.

[31] Feng Siyuan, Burchfiel Ben, Albina Toffee, and Tedrake Russ. 2023. TRI's Robots Learn New Manipulation Skills in an Afternoon. Here's How. Accessed: May 20, 2024.

[32] Jean-Jacques E Slotine and Weiping Li. 1987. On the adaptive control of robot manipulators. *The international journal of robotics research* 6, 3 (1987), 49–59.

[33] Maximilian Speicher, Brian D Hall, and Michael Nebeling. 2019. What is mixed reality?. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–15.

[34] Sebastian Starke. 2021. BioIK Asset for Unity3D. Accessed: May 29, 2024.

[35] Sebastian Starke, Norman Hendrich, Sven Magg, and Jianwei Zhang. 2016. An efficient hybridization of genetic algorithms and particle swarm optimization for inverse kinematics. In *2016 IEEE International Conference on Robotics and Biomimetics (ROBIO)*. IEEE, 1782–1789.

[36] Ryo Suzuki, Adnan Karim, Tian Xia, Hooman Hedayati, and Nicolai Marquardt. 2022. Augmented reality and robotics: A survey and taxonomy for ar-enhanced human-robot interaction and robotic interfaces. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*.

1–33.

[37] Unity Technologies. 2022. ROS TCP Connector. Accessed: May 25, 2024.

[38] Unity Technologies. 2022. URDF Importer. Accessed: May 25, 2024.

[39] Michael Walker, Thao Phung, Tathagata Chakraborti, Tom Williams, and Daniel Szafir. 2023. Virtual, augmented, and mixed reality for human-robot interaction: A survey and virtual design element taxonomy. *ACM Transactions on Human-Robot Interaction* 12, 4 (2023), 1–39.

[40] Charles W Wampler. 1986. Manipulator inverse kinematic solutions based on vector formulations and damped least-squares methods. *IEEE Transactions on Systems, Man, and Cybernetics* 16, 1 (1986), 93–101.

[41] Atsushi Watanabe, Tetsushi Ikeda, Yoichi Morales, Kazuhiko Shinozawa, Takahiro Miyashita, and Norihiro Hagita. 2015. Communicating robotic navigational intentions. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 5763–5769.

[42] Tsuneo Yoshikawa. 1985. Manipulability of robotic mechanisms. *The international journal of Robotics Research* 4, 2 (1985), 3–9.

[43] James E Young, Min Xin, and Ehud Sharlin. 2007. Robot expressionism through cartooning. In *Proceedings of the ACM/IEEE international conference on Human-robot interaction*. 309–316.

[44] Tianhao Zhang, Zoe McCarthy, Owen Jow, Dennis Lee, Xi Chen, Ken Goldberg, and Pieter Abbeel. 2018. Deep imitation learning for complex manipulation tasks from virtual reality teleoperation. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 5628–5635.

[45] Yuke Zhu, Josiah Wong, Ajay Mandlekar, Roberto Martín-Martín, Abhishek Joshi, Soroush Nasiriany, and Yifeng Zhu. 2020. robosuite: A modular simulation framework and benchmark for robot learning. *arXiv preprint arXiv:2009.12293* (2020).