2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

Training Humans to Teach Robots: Large and Lasting Skill Gains

Yuqing Zhu, Endong Sun and Matthew Howard

Abstract— **Recent evidence has shown that, contrary to expectations, it is difficult for novices to teach robots tasks through learning from demonstration (LfD). Novices often struggle with understanding the relationship between robot states and actions, leading to suboptimal demonstrations. This paper introduces a framework that leverages *machine teaching* algorithms to *train novices* in a controlled, ideal environment where optimal control parameters are predefined. The training enables participants to internalise fundamental control principles, preparing them to adapt to new skills that share similar properties. The study evaluates whether such teaching ability is (i) *retained* beyond the training period (including a long-term follow-up) and (ii) *generalised* so that novices teach robots more effectively in environments where control parameters are not predefined. It reports a series of between-subjects studies that demonstrate that trained novice teachers achieve a 75% improvement in teaching ability, with these gains retained even after guidance is removed, and exhibit a 71% enhancement in applying skills beyond the training content.**

I. Introduction

Robot learning from demonstration (LfD) is a technology that enables robots to learn tasks by observing and imitating human actions, eliminating the need for explicit programming. This approach greatly expands accessibility, allowing non-experts to train robots, and makes their deployment much simpler in a wide range of fields [1], [2]. LfD stands out for its ability to facilitate the learning of complex tasks efficiently, becoming a crucial direction in the advancement of modern robotics [3].

The human teacher's role in LfD is critical, as the quality of demonstrations directly influences the robot's learning and execution of tasks. Inexperienced teachers often provide ineffective demonstrations, leading to suboptimal robot performance. This is partly due to misconceptions about how robot states and actions relate [4]. For example, novices may intuitively believe that simply moving a robot towards a target position is sufficient, without realising that the dynamics of force, velocity, and position must be balanced to avoid overshooting or instability. Such misunderstandings limit their ability to provide high-quality demonstrations in tasks that involve complex state-action feedback. As variability and inconsistency in demonstrations can significantly degrade learning outcomes, this motivates use of structured support for novice teachers [5].

Recent efforts have focused on training human teachers to deliver higher-quality demonstrations, thereby improving LfD outcomes and teaching capabilities. For example, Sena and Howard [6] identify key teaching behaviours and provide feedback using optimal demonstration sets. However, their feedback methods rely on *expert-selected demonstrations*, a factor that limits its scalability to real-world scenarios where such expert advice may be unavailable. Other work has explored adaptive feedback methods to enhance human teaching in LfD. Gu, Croft, and Kulic [7] present a demonstration-based, post-hoc explainable AI system that adaptively selects trajectories to aid human understanding. However, in high-dimensional state–action spaces, the vast trajectory set makes selecting representative examples computationally expensive, and post-hoc summaries risk omitting key policy behaviours. Similarly, Heitkamp, Krieger, Friedman, *et al.* [8] presents an reinforcement learning-based scaffolding strategy for motor skill learning that performs well on simple tasks but struggles to generalise to complex, dynamic skills.

To address these limitations, this paper introduces a training framework that integrates machine teaching (MT) [9] to enhance teaching ability in LfD (see Fig. 1). Rather than proposing a new MT algorithm, MT is used to generate scalable training signals that help novices develop more effective demonstration strategies. The approach provides structured, real-time guidance in an ideal training environment with predefined optimal control parameters, enabling novices to internalise control principles. Feedback is delivered in an intuitive visual form, reducing cognitive load while still conveying essential corrections. Crucially, the framework is designed to ensure that the *knowledge gained generalises to new, unseen skills* without requiring further guidance, leading to improvements that are both *retained after training (including in a long-term follow-up)* and *transferable across skill types*.

The contributions of this paper are to (i) propose a scalable training framework that *automatically-generates* optimal guidance to help novices develop control strategies for robotic skills, and (ii) verify the resultant improvement of teaching *quality*, its *retention* beyond the training period (with evidence of durability over a one-year interval), and its *generalisation* to new motor skills. Three between-subjects studies are reported, in which novice teachers are asked to teach motor skills of varying complexity to a robot. The results demonstrate that subjects who receive training show *>70% improvement in teaching ability*, on average, *even for skills not seen in training* and this improvement is retained for *at least a year after training*.

This work was supported by King's College London and the China Scholarship Council.

Yuqing Zhu (Corresponding author: yuqing.zhu@kcl.ac.uk), Endong Sun and Matthew Howard are with the Centre for Robotics Research, Department of Engineering, King's College London, UK.

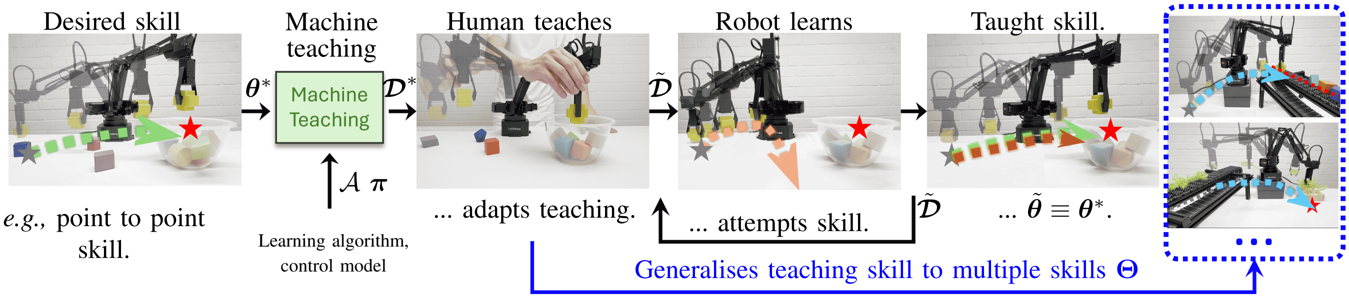

Fig. 1: The overview of framework: the framework uses MT to generate optimal guidance (green box), helping novices refine their demonstrations. Initially, suboptimal demonstrations lead to failed learning (light orange), but with training, teachers improve their teaching ability to provide high-quality demonstrations, and robots execute skills well (dark orange). This teaching ability can be retained and further generalised to unseen skills (blue box).

II. Problem Definition

In the problems considered in this paper, motor skills are represented as a control policy in the form

$$\boldsymbol{u} = \boldsymbol{\pi}(\boldsymbol{x}, \boldsymbol{\theta}) \tag{1}$$

where $\boldsymbol{\pi}(\cdot)$ maps the robot's state, $\boldsymbol{x} \in \mathbb{R}^{\mathcal{P}}$, to the corresponding action, $\boldsymbol{u} \in \mathbb{R}^{\mathcal{R}}$. The variable $\boldsymbol{\theta} \in \mathbb{R}^{\mathcal{S}}$ denotes the skill parameters.

Skills are communicated to the robot through teaching, *i.e.,* by providing a set of demonstration data consisting of a sequence of state-action tuples $\boldsymbol{\mathcal{D}} = \{(\boldsymbol{x}_1, \boldsymbol{u}_1) ..., (\boldsymbol{x}_{\mathcal{N}}, \boldsymbol{u}_{\mathcal{N}})\} \in (\mathbb{R}^{\mathcal{P} \times \mathcal{N}}, \mathbb{R}^{\mathcal{R} \times \mathcal{N}})$. The robot learner learns the skill parameters from $\boldsymbol{\mathcal{D}}$ with a learning algorithm

$$\tilde{\boldsymbol{\theta}} = \mathcal{A}(\boldsymbol{\mathcal{D}}) \tag{2}$$

leading to the learnt skill

$$\tilde{\boldsymbol{u}} = \boldsymbol{\pi}(\boldsymbol{x}, \tilde{\boldsymbol{\theta}}). \tag{3}$$

Note that, the success of learning depends heavily on the *quality and quantity of the demonstration data*, something that can not easily be assured when the data is provided by novice human teachers.

One approach to deriving high-quality data is to apply the principles of machine teaching. In MT, the aim is to find the *optimal data* that will lead the learner, to learn the target model $\boldsymbol{\theta}^*$. One formulation of this is as a bi-level optimisation problem [10]

$$\boldsymbol{\mathcal{D}}^* = \arg\min_{\boldsymbol{\mathcal{D}} \in \mathfrak{D}} \rho(\tilde{\boldsymbol{\theta}}, \boldsymbol{\theta}^*) \tag{4}$$

$$\text{s.t.} \quad \tilde{\boldsymbol{\theta}} = \mathcal{A}(\boldsymbol{\mathcal{D}}) \quad \text{and} \quad \varepsilon(\boldsymbol{\mathcal{D}}) \leq \varepsilon_{max}. \tag{5}$$

The teaching risk $\rho(\cdot)$ quantifies how accurately the target model can be learned using data from the space of possible data sets $\mathfrak{D}$ and $\varepsilon(\cdot)$ represents the teaching effort (which, in this formulation, is limited to a maximum of ε_{max}).

The solution of (4)-(5) requires (i) the target model $\boldsymbol{\theta}^*$ to be explicitly known, and (ii) the optimisation to be tractable. In many real-world applications this is difficult to guarantee, particularly when dealing with complex robotic motor skills. However, *human teachers' implicit knowledge can play a crucial role in overcoming these limitations*. Specifically, this study investigates whether MT can shape novice teachers' strategies in ways that align with optimal principles, even for skills where *explicit solutions to (4)-(5) are unavailable*. The expectation is that by training under conditions where the ideal parameters $\boldsymbol{\theta}^*$ and data $\boldsymbol{\mathcal{D}}^*$ are known, novice teachers will internalise the relationship between the two, enabling them to generalise their teaching ability to new, unseen skills. This approach leverages the *optimality benefits of MT* while harnessing the *versatility and adaptability of human teachers* to transfer learning across skills. In this paper, optimal actions are required only for the controlled training tasks. The aim is to induce teaching behaviours that remain useful when teaching downstream skills where such optimal actions are not available. For longer-horizon or sequential tasks, a practical extension is to apply guidance along representative sub-goals or sub-skills and then rely on the learned teaching behaviour when demonstrating the full task.

In summary, the hypotheses evaluated in this study are:

- h_1 Teaching ability improves when novice teachers receive guidance derived from the principles of MT.
- h_2 After guidance is removed, the improved teaching ability is retained.
- h_3 The improved teaching ability learnt in one skill can be transferred to another skill.

The next section describes the materials and methods used to test these.

III. Training Framework

A. MT Formulation

The core to the proposed framework is to derive guidance from MT as a training signal for novice human teachers. For this, the following modelling assumptions and design choices are made.

1) Modelling the Learner: To balance the demands of simplicity, robustness and scalability, in this paper, all robot learners use control policies of the form

$$u = \boldsymbol{\theta}^{\top} \boldsymbol{\phi}(\boldsymbol{x}) \tag{6}$$

where $\boldsymbol{\phi}(\boldsymbol{x}) \in \mathbb{R}^{\mathcal{S}}$ is a mapping from the state to a suitable feature space. Note that, depending on the choice of features, this model may represent policies of arbitrary complexity, and enables the inclusion of robot-specific features (such as kinematic mappings, see §IV). [1]

The parameter estimate $\tilde{\boldsymbol{\theta}}$ is found through ridge regression for which the closed-form solution is

$$\tilde{\boldsymbol{\theta}} = (\boldsymbol{\Phi}\boldsymbol{\Phi}^\top + \lambda \boldsymbol{I})^{-1}\boldsymbol{\Phi}^\top \mathbf{u} \tag{7}$$

where $\mathbf{u} = (u_1, ..., u_\mathcal{N})$, the nth column of $\boldsymbol{\Phi} \in \mathbb{R}^{\mathcal{S}\times\mathcal{N}}$ contains $\boldsymbol{\phi}(\boldsymbol{x}_n)$ and λ is a regularisation parameter. Assuming the robot learns through (7), the specification of the MT problem requires that appropriate teaching risk and effort functions are selected. A linear learner is adopted here to obtain an unambiguous, tractable mapping from demonstrations to guidance.

2) Teaching Effort: In this study, the *number of demonstrations given by the teacher* is used as a parsimonious, platform-independent measure of effort *i.e.,*

$$\varepsilon(\boldsymbol{\mathcal{D}}) = \mathcal{N}. \tag{8}$$

Note that, the formulation (4)-(5) imposes an upper bound on the teaching effort. Considering (7), the minimal number of data points required for effective learning corresponds to the number of data points required to span the feature space [11], so $\varepsilon_{max} = \mathcal{S}$ is used here.

3) Teaching Risk: In this study, the ℓ^2-norm between the learnt and target parameters is used as the teaching risk, *i.e.,*

$$\rho(\tilde{\boldsymbol{\theta}}, \boldsymbol{\theta}^*) = \|\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|_2. \tag{9}$$

This provides a direct measure as to whether teaching results in learnt parameters that are close to the target, making it suitable for use in the training environment (where $\boldsymbol{\theta}^*$ is explicitly known).

B. MT-based Training

The training framework in this study uses direct guidance on the quality of demonstrations to shape novices' teaching behaviour. For this, guidance is derived from an analysis on how demonstrations affect the learnt parameter based on MT. Actions are elicited at pre-selected query states to standardise the state distribution across participants and provide an immediate per-state guidance signal.

1) Selection of States: To scaffold learning, states are pre-selected for the novices. This approach is supported by recent work [6], which shows that humans often struggle to select states that effectively test a robot's ability to generalise. Additionally, it has been shown that for the learning algorithm (7), the optimal choice of states is non-unique [9], a fact that causes confusion for non-technical novice teachers.

In the experiments reported in this study, states are selected via a pseudo-random process that uniformly samples from the task-relevant state space region. For example, in the character-writing task this corresponds to sampling across the drawing canvas/workspace rather than only at the trajectory start. States are then iteratively added to a set of *query states* $\{\boldsymbol{x}_1, ..., \boldsymbol{x}_\mathcal{N}\}$, subject to the constraint that each new state does not cause $\boldsymbol{\Phi}$ to become ill-conditioned.

[1]For simplicity, the following describes learning in the case that $\mathcal{R} = 1$, this may be trivially extended to $\mathcal{R} > 1$ by replacing $\boldsymbol{\theta}^*$ with the appropriate skill parameter matrix $\boldsymbol{\Theta}$.

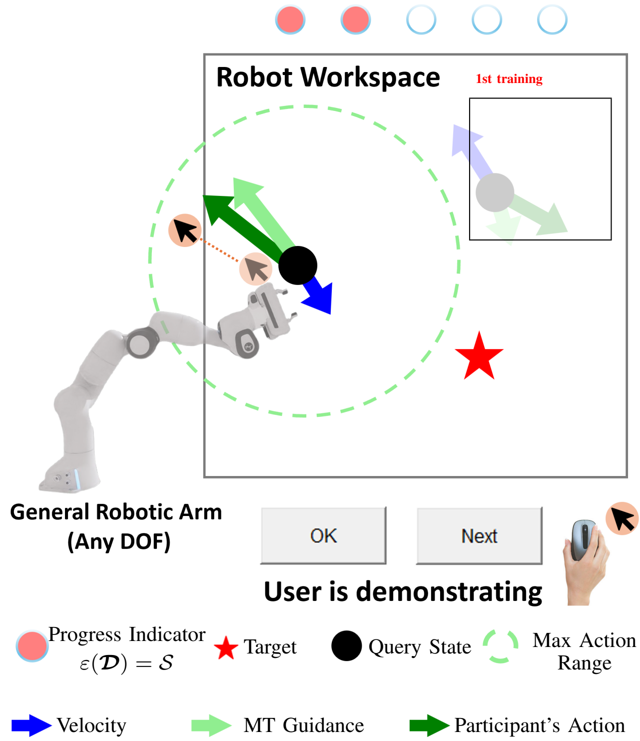

Fig. 2: Simulation platform user interface (UI) show during training for Experiment 1 (see §IV). Novice teachers can manipulate the arm by dragging the mouse to specify actions (dark green arrow) corresponding to the given keyframe states (black dot, blue arrow). MT-based feedback is provided through the optimal action (light green arrow).

2) Selection of Actions: Given the query states, novice teachers are then required to select the appropriate action u_n for the corresponding query state $\boldsymbol{x}_n$. This is repeated for all query states to complete the data set and demonstrate the skill. For this, the MT risk function is used to guide their choice, as follows.

Substituting (7) into (9), and noting that $\boldsymbol{\theta}^*$ is independent of $\boldsymbol{\Phi}$ and $\mathbf{u}$

$$\rho(\tilde{\boldsymbol{\theta}}, \boldsymbol{\theta}^*) = \left\| \left(\boldsymbol{\Phi}\boldsymbol{\Phi}^\top + \lambda \boldsymbol{I}\right)^{-1} \boldsymbol{\Phi}(\tilde{\mathbf{u}} - \mathbf{u}^*)^\top \right\|_2 \tag{10}$$

where $\mathbf{u}^* = (u_1^*, ..., u_\mathcal{N}^*)$ are the *optimal* action demonstrations (*i.e.,* those that would lead to $\tilde{\boldsymbol{\theta}} = \boldsymbol{\theta}^*$ when learnt via (7)).

Using the submultiplicative property of matrix norms [12], this can be written as

$$\rho(\tilde{\boldsymbol{\theta}}, \boldsymbol{\theta}^*) \leq \underbrace{\left\| \left(\boldsymbol{\Phi}^\top\boldsymbol{\Phi} + \lambda \boldsymbol{I}\right)^{-1} \boldsymbol{\Phi}^\top \right\|_2}_{\rho_1} \underbrace{\|\tilde{\mathbf{u}} - \mathbf{u}^*\|_2^\top}_{\rho_2}. \tag{11}$$

Here, ρ_2 captures the difference between the actions provided during LfD and the desired actions. This provides a clear, quantifiable measure of demonstration quality, allowing for immediate evaluation of how the demonstrations can be improved, and therefore forms the basis of the training guidance.

C. Guidance Design

The key considerations used to develop the proposed guidance system are that the feedback must (i) convey all critical pieces of information (*e.g.,* how much effort to expend, the direction and magnitude of the actions), (ii) be intuitive, enabling the novice teacher to interpret without requiring extensive technical knowledge and (iii) be provided at the right moment (too early may disrupt the demonstration, while too late risks the teacher forgetting details).

Based on these considerations, immediate post-demonstration visual feedback is used (see Fig. 2 and Fig. 3). The underlying demonstration interface remains the same across conditions, and only the presence or absence of the guidance overlay differs. The feedback consists of (i) a depiction of the robot (or the robot itself), (ii) arrows representing the direction and magnitude of the correct action $\mathbf{u}^*$ and that provided by the participant $\tilde{\mathbf{u}}$, (iii) a progress indicator to track teaching effort that advances after each demonstration and (iv) state information (*e.g.,* the position of the robot's end effector, arrows representing velocity). States are shown sequentially, with one state appearing after another, to ensure clarity and focus for the participant. The displayed states are selected from the portion of the movement that is most relevant to the task objective.

[2] This design is general and flexible, making it applicable to a wide range of robotic platforms, skills and environments.

IV. Evaluation

In this section, experiments[3] are reported in which the effectiveness of the proposed training framework is assessed in (i) a simulated platform and (ii) a physical robot.

Experiment 1. Dynamic Control

In this experiment, a simulation of a two-link robotic arm is used to assess novice teachers' ability to teach force-controlled skills.

1) Robot setup: The robot's state is represented as $\boldsymbol{x} = (q_1, q_2, \dot{q}_1, \dot{q}_2)^\top$, where $\boldsymbol{q} = (q_1, q_2)^\top$ and $\dot{\boldsymbol{q}} = (\dot{q}_1, \dot{q}_2)^\top$ denote the position and velocity of the robot's joints, respectively. The actions consist of the force in end-effector space $\boldsymbol{u} = (f_1, f_2)^\top$. Each link of the arm has length $l = 1\,\mathrm{m}$ and mass $m = 1\,\mathrm{kg}$. Gravity g is $9.81\,\mathrm{m\,s^{-2}}$. The arm operates without force limits, and has a sampling rate of $1\,\mathrm{kHz}$.

The dynamics of the robotic arm are modelled as

$$\boldsymbol{M}(\boldsymbol{q})\ddot{\boldsymbol{q}} + \boldsymbol{C}(\boldsymbol{q}, \dot{\boldsymbol{q}})\dot{\boldsymbol{q}} + \boldsymbol{g}(\boldsymbol{q}) = \boldsymbol{J}^\top(\boldsymbol{q})\boldsymbol{u} \tag{12}$$

where $\ddot{\boldsymbol{q}}$ is the joint angular acceleration, $\boldsymbol{M} \in \mathbb{R}^{2\times 2}$ is the mass matrix, $\boldsymbol{C} \in \mathbb{R}^{2\times 2}$ represents Coriolis and centrifugal forces and $\boldsymbol{g} \in \mathbb{R}^{2\times 1}$ is the gravity vector. The applied $\boldsymbol{u}$ is converted into joint torques via the Jacobian $\boldsymbol{J} \in \mathbb{R}^{2\times 2}$.

2) Motor skills: The class of skills that must be taught to the robot consists of *convergent motion* from any point in the workspace to a target. Specifically, the two skills studied here are to reach to a *target point* ($\boldsymbol{\Theta}_1$) , and a *target line* ($\boldsymbol{\Theta}_2$) . These skills can be represented by a coupled pair of controllers of the form (6) with $\boldsymbol{\phi}(\boldsymbol{x}) = (r_1, r_2, \dot{r}_1, \dot{r}_2, 1)^\top$ where $\boldsymbol{r} = (r_1, r_2)^\top$ and $\dot{\boldsymbol{r}} = (\dot{r}_1, \dot{r}_2)^\top$ denote the position and velocity of the robot's end-effector, respectively, and

$$\boldsymbol{\Theta}_i = \begin{bmatrix} k_1 & k_2 & d_1 & d_2 & r_1^* \\ k_3 & k_4 & d_3 & d_4 & r_2^* \end{bmatrix}^\top, \quad i \in \{1, 2\} \tag{13}$$

with $\mathrm{vec}\,(\boldsymbol{\Theta}_1) = (-1, 0, -1, 0, 0.8, 0, -1, 0, -1, 1.2)^\top$ and $\mathrm{vec}\,(\boldsymbol{\Theta}_2) = (-1, 0, -1, 0, 0.8, 0, 0, 0, -1, 0)^\top$.

3) Experiment Protocol: The experiments follow a between-subjects design, with target and control groups asked to teach the motor skills across a series of phases. Only the target group receives guidance during the training phase (P3–P10) to measure the effectiveness of MT guidance against the teaching strategy of the control group (h_1). Performance in the pre-training phases (P1, P2) and post-training phases (P11, P12) is compared to assess retention (h_2) and generalisation (h_3) of the teaching ability. Specifically, the experimental phases are

- P1 *Skill 1, no guidance.* Participants give demonstrations to teach $\boldsymbol{\Theta}_1$ without guidance.
- P2 *Skill 2, no guidance.* Participants give demonstrations to teach $\boldsymbol{\Theta}_2$ without guidance.
- P3 *Skill 1, with guidance.* Participants give demonstrations to teach $\boldsymbol{\Theta}_1$. The target group receives guidance, while the control group does not.
- P4-10 Participants repeat P3.
- P11 Participants repeat P1.
- P12 Participants repeat P2.

To determine statistical significance, paired t-tests and ANOVA are applied to analyse the impact of MT-based guidance across these conditions. Across all phases, participants received the same standardised task instructions, and the experimenter provided no additional coaching or performance feedback during data collection.

4) Participants: The sample size for the experiment is determined using *G*Power 3.1* [13]. A one-tailed t-test, with an effect size of 0.9, an alpha level of 0.05, and a statistical power of 0.8 indicates a required sample size of $n = 32$ participants. Participants are randomly split into groups of equal size. A single-blind procedure is used to ensure that participants are unaware of their group assignment. The criterion for inclusion into the cohort is that the participant has no prior experience working with robots.

5) Evaluation Metrics: Subjects' teaching ability is assessed in terms of the robot's learning outcomes. Specifically, robot learning outcomes are measured using the ℓ_2-norm

$$E_{\ell_2} = \sqrt{(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)^\top(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)} / \sqrt{\boldsymbol{\theta}^{*\top}\boldsymbol{\theta}^*},\ \boldsymbol{\theta}^* = \mathrm{vec}(\boldsymbol{\Theta}). \tag{14}$$

[2]The video is submitted as supplementary material.

[3]This study is conducted under the approval of the King's College London Research Ethics Committee, Ref.: MRPP-22/23-37844. Informed consent was obtained from all experimental participants.

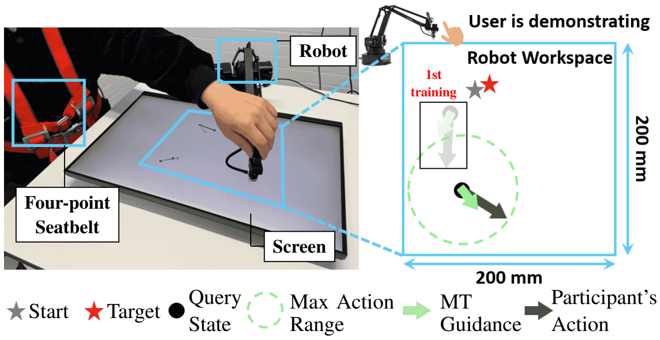

Fig. 3: Physical platform UI consisting of a uArm robot and display screen. A four-point seat belt ensures the participant maintains a consistent viewpoint.

This provides a benchmark for how closely the parameters learnt by the robot match the target parameters. In the controlled training setting, where the target parameters are known by design, this normalised parameter error provides a direct and comparable measure of teaching effectiveness across skills and experimental phases.

Experiment 2. Reaching with a Physical Robot

The aim of this experiment is to test the teaching of simple reaching skills with a physical robot.

1) Robot setup: The robotic platform chosen for this experiment is the uArm Swift Pro. This is a 4-degree-of-freedom robotic arm with two $0.15\,\mathrm{m}$ links and a total weight of $2.2\,\mathrm{kg}$. It operates within a working range of $50\,\mathrm{mm}$ to $320\,\mathrm{mm}$ and has maximum speed of $100\,\mathrm{mm\,s^{-1}}$. With stepper motors and 12-bit encoders, it achieves a positional repeatability of $0.2\,\mathrm{mm}$, ensuring high precision in training skills. More detailed technical specifications can be found in the manufacturer's documentation [14].

The robot is kinematically controlled, with its state defined by the instantaneous position of its joints $\boldsymbol{x} = (q_1, q_2, q_3, q_4)^\top$ and the action given by $\boldsymbol{u} = (\Delta r_1, \Delta r_2)^\top$, where Δr_1 and Δr_2 represent the desired change in the end-effector position. Note that, the uArm Swift Pro, like many industrial robots is designed to prioritise positional accuracy in its control.

2) Motor skills: Similar to Experiment 1, the motor skills to be taught represent *convergent motion to a target point* ($\boldsymbol{\Theta}_3$) , and a *target line* ($\boldsymbol{\Theta}_4$) defined by a pair of coupled controllers with parameters

$$\boldsymbol{\Theta}_i = \begin{bmatrix} k_1 & k_2 & r_1^* \\ k_3 & k_4 & r_2^* \end{bmatrix}^\top, \quad i \in \{3, 4\} \tag{15}$$

with the corresponding feature vector $\boldsymbol{\phi}(\boldsymbol{x}) = (r_1, r_2, 1)^\top$. Here, $\mathrm{vec}\,(\boldsymbol{\Theta}_3) = (-0.2, 0, 0.46, 0, -0.2, 0.22)^\top$ and $\mathrm{vec}\,(\boldsymbol{\Theta}_4) = (-0.04, 0.08, 0, 0.08, -0.16, 0)^\top$.

3) Experiment Protocol and Evaluation Metrics: The same multiple-phase protocol (P1–P12) and evaluation metrics as in Experiment 1 are applied here for consistency.

4) Participants: A new set of $n = 32$ participants is recruited using the same *G*Power 3.1* setting and selection criteria as in Experiment 1. Eligible participants are randomly assigned to the target or control group.

Experiment 3. Teaching Hand-written Characters

The aim of this experiment is to test MT-based framework when the target skills involve complex trajectories that must be represented with more general function approximators.

1) Robot setup: This experiment also employs the uArm Swift Pro with the same state and action representation as in Experiment 2, but introduces complex control through policies based on radial basis function networks (RBFNs). Specifically, the control policy takes the form (6) where the feature vector contains elements

$$\phi_p(\boldsymbol{x}) = \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{c}_p\|^2}{2\sigma^2}\right) \tag{16}$$

where $\sigma = 0.53$ is the width of radial basis functions (RBFs) and $\boldsymbol{c}_p \in \mathbb{R}^2$ represents the centre of the pth RBF. In this study, four RBFs are employed, with their centres located on a grid at the positions $(20, 20), (20, 180), (180, 20), (180, 180)$.

2) Motor skills: In this experiment, the target motor skills are to draw a *letter 'O'* ($\boldsymbol{\Theta}_5$) , and to draw a *letter 'S'* ($\boldsymbol{\Theta}_6$) . The target parameters are set by training the RBFN on a canonical data set containing hand-written samples of the letters.

3) Experiment Protocol and Evaluation metrics: This experiment follows the same protocol (P1–P12) and uses the same evaluation metrics as the preceding experiments.

4) Participants: In this experiment, $n = 16$ participants are recruited, randomly split into equal-sized target and control groups. The eligibility criteria remain the same as in previous experiments.

A. Results

To check for possible confounding effects, such as inherent differences in skill complexity or participant bias, a post-hoc analysis of data comparing P1 and P2 across all three experiments was conducted. This confirmed that, prior to training, there is no significant difference in teaching ability observed (i) between the two skills, or (ii) between the target and control groups ($p > 0.05$). This validates that any observed improvements in later phases are due to the training intervention. Moreover, since the control group receives the same task exposure and interface practice across phases but does not show comparable reductions in error, the observed gains are consistent with an effect of the training guidance rather than practice alone.

1) Effectiveness of Machine Teaching Guidance (h_1): The results of the t-tests indicate that h_1 is supported, showing that machine teaching guidance significantly improves novice teaching ability. In Experiment 1 (see Fig. 4(a-1)), the comparison between P3 and P10 reveals a 65.3% reduction in teaching error for the target group, compared to no reduction in error for the control group. The results

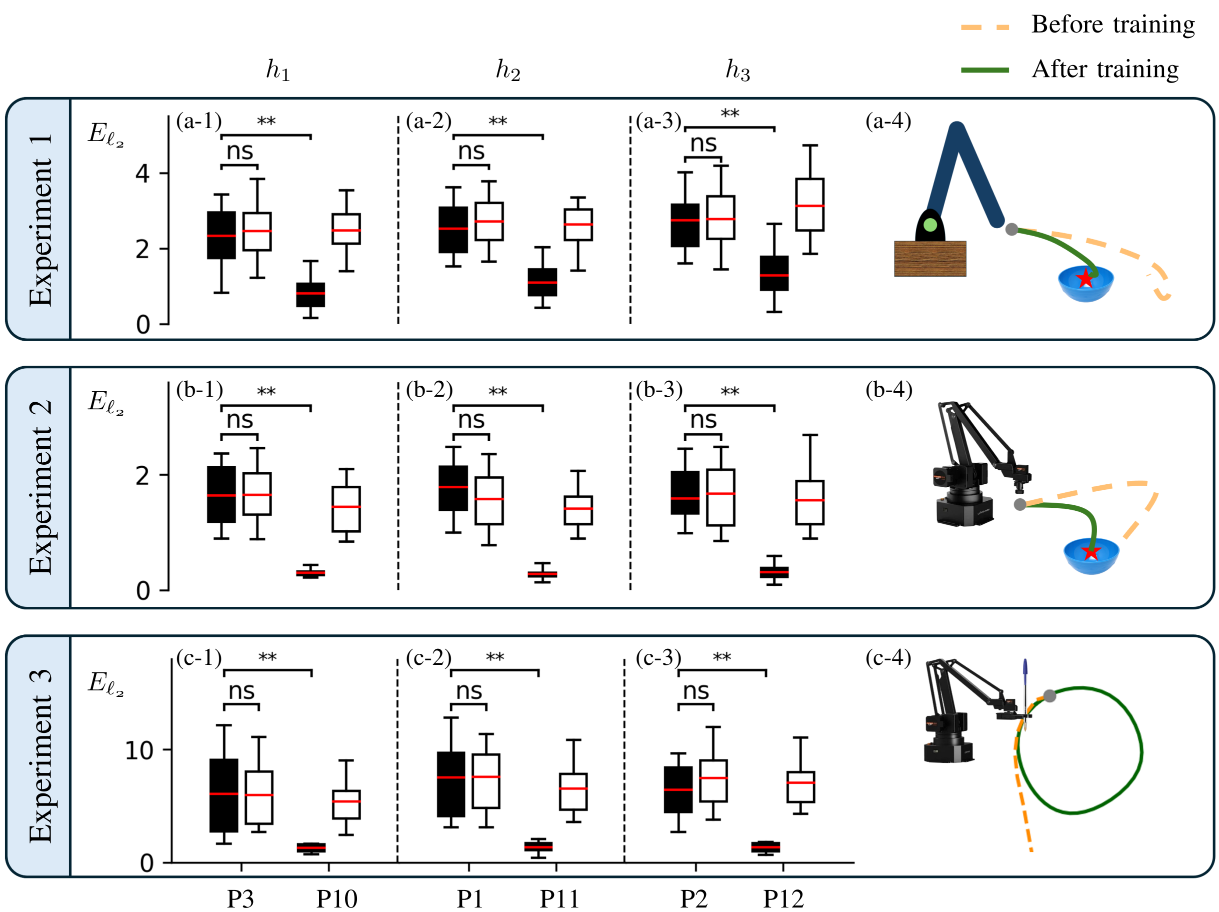

Fig. 4: Results of Experiment 1-Experiment 3. Subjects' performance in teaching skills: (a-1/2/3, b-1/2/3, c-1/2/3) E_{ℓ_2} of the target (black) and control (white) groups in different phase comparisons: P3 vs. P10, P1 vs. P11, and P2 vs. P12, for Experiment 1, Experiment 2 and Experiment 3 respectively. Red bars indicate the mean. (a-4, b-4, c-4) Example trajectories from a representative target group subject in each experiment, showing robot motions before and after training.

of Experiment 2 follow a similar pattern (see Fig. 4(b-1)), with the target group achieving an 82.3% reduction in error compared to 12.5% in the control group. In Experiment 3, the teaching error reduces by 79.3% for the target group, compared to 9.4% in the control group (see Fig. 4(c-1)). Taken together, these findings indicate that structured guidance enables participants to adapt their teaching strategies effectively, *leading to improved robot learning outcomes*. Furthermore, the target group's teaching error becomes more *consistent*, as indicated by the reduction in the range and interquartile range (whiskers and boxes, respectively).

2) Retention of Teaching Ability (h_2): The results of the t-tests support h_2, confirming that the benefits of guided teaching extend beyond the training phase. To evaluate how teaching performance persists once guidance is removed, the target group's performance in the final guided phase is compared with their performance in the first unguided phase (P1 vs. P11). As illustrated in Fig. 4(a-2), participants show no observable drop in performance after guidance is withdrawn, indicating that the improvements achieved during training are retained. The target group sustains a reduction in teaching error (56.8%) relative to their initial attempt, with

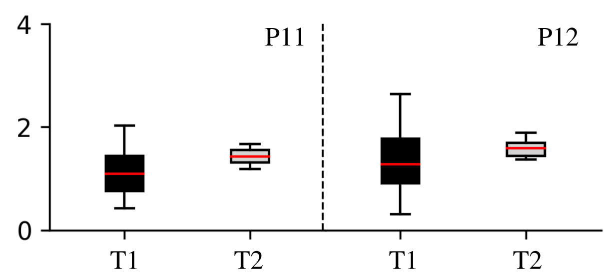

Fig. 5: Long-term retention performance in Experiment 1: initial test (T1, black) and one-year-later retest (T2, gray). Results are shown separately for training skill (left) and unseen skill (right). Performance remains stable across sessions, indicating strong retention.

performance in the unguided phase remaining at the same level as in the guided phase. Similar patterns are observed in Experiment 2 (Fig. 4(b-2)), where the target group maintains an 84.7% improvement after guidance is withdrawn, and in Experiment 3 (Fig. 4(c-2)), where an 82.5% improvement is stably retained.

To evaluate long-term retention of teaching skill, partici-

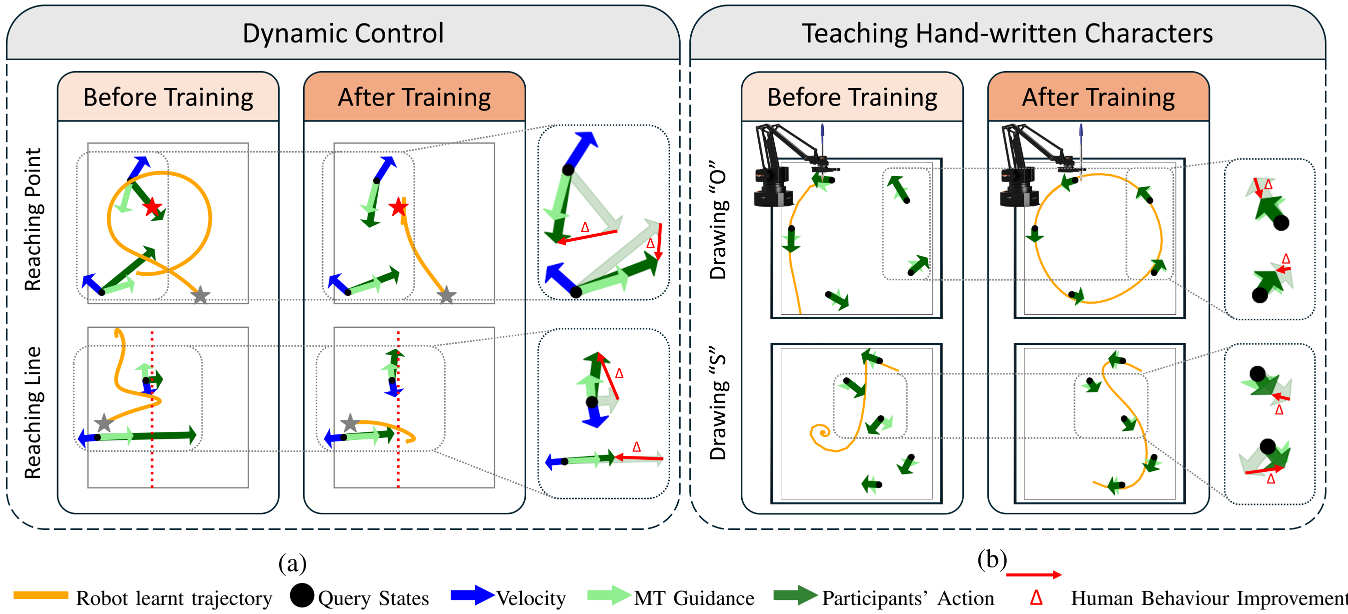

Fig. 6: Evolution of teaching behaviour before and after training (across phases P1, P2, P11, P12) in (a) Experiment 1: dynamic control, and (b) Experiment 3: teaching hand-written characters. Representative participants' demonstrations and the resulting robot trajectories are shown.

pants were contacted approximately one year after training. Of the initial target group, n=3 responded to the request to participate in a follow-up study. As shown in Fig. 5, this group retained their original performance gains, indicating *durable improvements without additional practice*. These results indicate that structured guidance produces teaching behaviours that carry over immediately into the unguided phase without measurable degradation in performance and can be sustained over extended periods. While this reflects both short-term and long-term retention, the findings demonstrate that such guidance yields strategies that remain effective well beyond its removal.

3) Generalisation to New Skills (h_3): The results of the t-tests support h_3, indicating that improvements in teaching ability extend beyond the originally trained skills. The transferability of teaching ability to new skills is evaluated by comparing performance metrics between P2 and P12. In Experiment 1 (see Fig. 4(a-3)), the target group's teaching error for teaching $\mathbf{\Theta}_2$ reduces by 53.7% post-training, compared to 8.7% in the control group.

This finding is reproduced in Experiment 2 (see Fig. 4(b-3)) with an 80.5% reduction for the target group, compared to 6.7% for the control group, and again in Experiment 3 (Fig. 4(c-3)) with 79.2% reduction in teaching error in the target group compared to 5.5% in the control group. These results imply that MT-based guidance not only enhances teaching ability for the skills seen in training, but also *transfers to the teaching of other related skills for which no prior training is given.*

4) Teaching Strategy Analysis: The improvements in teaching ability can be better understood from an analysis of the change in teaching behaviour across phases. Fig. 6 shows the demonstrations and resultant learning outcomes (robot trajectories), for a representative subject from the target groups in Experiment 1 and Experiment 3. As can be seen, before training, participants often applied actions of excessive magnitude and poorly aligned direction, typically pushing strongly toward the contour without correcting for the underlying velocity of the state (blue arrows). This mismatch led to distorted trajectories, and unstable or incomplete shapes (*e.g.,* spirals instead of smooth "S", or collapsed "O"). After training, participants' actions became more precise, with magnitudes better scaled and directions closely aligned to the motion teacher's guidance. This resulted in smooth and stable reproduction of the intended trajectories, reflecting a clear improvement in both consistency and accuracy of teaching behaviour.

V. Conclusions

This study proposes an MT-based training framework that assists novice teachers in providing high-quality demonstrations of dynamic skills to robots. The experimental results clearly indicate that MT-based guidance significantly enhances teaching ability, retention and transferability. On average, the participants in the target group exhibit an immediate reduction in teaching error of approximately 75%, an improvement that is retained even when guidance is removed. Moreover, the average improvement in teaching error when teaching skills not seen in training is 71%, showing transferability. Trajectory analyses further reveal that the framework promotes more deliberate and effective teaching strategies, while control participants continue to rely on less efficient approaches.

An important observation across experiments is the difference in improvement between kinematic and dynamic skills. Although the kinematic tasks in Experiment 3 involve seemingly more complex trajectories, participants achieved greater gains compared to the dynamic force-control tasks in Experiment 1. This suggests that direct manipulation of end-effector positions in kinematic control is more intuitive for human teachers, whereas dynamic force control is more challenging due to its sensitivity to motion dynamics.

Future work will scale evaluation to more realistic settings, such as industrial environments and assembly tasks. MT-guided teaching will also be combined with learning from failed demonstrations, which should improve out-of-distribution generalisation and failure recovery. These steps will extend the scope of the framework while building on the foundation established here. Future work will also compare against alternative training baselines (*e.g.,* video-only instruction), evaluate transfer when teaching via full kinesthetic demonstrations, and extend the approach to longer-horizon sequential skills.

REFERENCES

[1] A. Bechar and C. Vigneault, "Agricultural robots for field operations: Concepts and components," *Biosystems Engineering*, vol. 149, pp. 94–111, 2017.

[2] M. Bojarski, D. Del Testa, D. Dworakowski, B. Firner, B. Flepp, P. Goyal, and X. Zhang, "End to end learning for self-driving cars," *arXiv preprint arXiv:1604.07316*, 2016.

[3] C. Breazeal, K. Dautenhahn, and T. Kanda, "Social robotics," in *Springer Handbook of Robotics*, Springer, 2016, pp. 1935–1972.

[4] C. G. Atkeson and S. Schaal, "Robot learning from demonstration," in *Proceedings of the Fourteenth International Conference on Machine Learning*, vol. 97, 1997, pp. 12–20.

[5] M. Cakmak and A. L. Thomaz, "Designing robot learners that ask good questions," in *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, 2012, pp. 17–24.

[6] A. Sena and M. Howard, "Quantifying teaching behavior in robot learning from demonstration," *IJRR*, vol. 39, no. 1, pp. 54–72, 2020.

[7] M. Gu, E. Croft, and D. Kulic, "Demonstration based explainable ai for learning from demonstration methods," *arXiv preprint arXiv:2410.05715*, 2024.

[8] D. Heitkamp, K. Krieger, J. Friedman, and A. Moringen, "Optimizing a scaffold to guide motor skill learning," in *AAAI-22 Workshop on Reinforcement Learning for Education: Opportunities and Challenges*, 2022.

[9] X. Zhu, "Machine teaching: An inverse problem to machine learning and an approach toward optimal education," in *AAAI*, 2015.

[10] X. Zhu, A. Singla, S. Zilles, and A. N. Rafferty, "An overview of machine teaching," *arXiv preprint arXiv:1801.05927*, 2018.

[11] J. Liu and X. Zhu, "The teaching dimension of linear learners," *Journal of Machine Learning Research*, vol. 17, no. 162, pp. 1–25, 2016.

[12] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge university press, 2012.

[13] F. Faul, E. Erdfelder, A. Buchner, and A.-G. Lang, "Statistical power analyses using g* power 3.1: Tests for correlation and regression analyses," *Behavior research methods*, vol. 41, no. 4, pp. 1149–1160, 2009.

[14] uFactory, *Uarm swift pro developer guide v1.0.6*, Available online: `http://ufactory.cc`, 2018.