

LOPAL: Local Performance-Aware Active Learning From Imperfect Demonstrations

Johannes Heidersberger^{ID}, Shail Jadav^{ID}, and Dongheui Lee^{ID}, *Senior Member, IEEE*

Abstract—Learning from Demonstration (LfD) enables intuitive robot skill acquisition by allowing robots to learn directly from human task demonstrations. However, current methods often fail to address the fact that due to suboptimal and inconsistent human behavior, the quality of the demonstration can vary within each demonstration. Therefore, we introduce LOPAL (Local Performance-aware Active Learning), an active learning approach that leverages this local demonstration quality information. Our approach consists of two synergistic components. First, a local performance-driven LfD method uses a Gaussian Mixture Model (GMM) to encode both the demonstrated trajectories and their associated local quality assessments. This enables the generation of trajectories that outperform the imperfect demonstrations by utilizing complementary local data of high performance. Second, active data acquisition allows to improve beyond the imperfect demonstrations by collecting additional informative samples. In areas missing good data, the user is actively requested to provide corrections through a shared autonomy (SA) mechanism, while the robot autonomously executes the learned behavior. The efficacy of LOPAL was validated in both a simulation and a real-world experiment. The results from a real-world pipe inspection task showed that the proposed approach can achieve up to 27.31% improvement in task performance while also reducing the effort required to collect the demonstrations.

Index Terms—Learning from demonstration, physical human-robot interaction, active learning, shared autonomy, imperfect demonstration.

I. INTRODUCTION

TRADITIONAL manual programming can be time-consuming and laborious. Learning from Demonstration

Received 25 November 2025; accepted 1 May 2026. Date of publication 29 May 2026; date of current version 17 June 2026. This article was recommended for publication by Associate Editor Fethiye Irmak Dogan and Editor Sonia Chernova upon evaluation of the reviewers' comments. This work was supported in part by Austrian Science Fund (FWF) under the project "LunarAssembly" under Grant PIN5145724, and in part by Vienna Science and Technology Fund (WWTF) under the project "SafeDiffusion" under Grant ICT25068. (*Corresponding author: Johannes Heidersberger.*)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by TU Wien Research Ethics Committee under Application No. 045_19012024_TUWREC.

Johannes Heidersberger and Shail Jadav are with the Autonomous Systems Lab, Institute of Computer Technology, TU Wien, 1040 Vienna, Austria (e-mail: johannes.heidersberger@tuwien.ac.at; shail.jadav@tuwien.ac.at).

Dongheui Lee is with the Autonomous Systems Lab, Institute of Computer Technology, TU Wien, 1040 Vienna, Austria, and also with the Institute of Robotics and Mechatronics, German Aerospace Center (DLR), 82234 Weßling, Germany (e-mail: dongheui.lee@tuwien.ac.at).

This article has supplementary downloadable material available at <https://doi.org/10.1109/LRA.2026.3698364>, provided by the authors.

Digital Object Identifier 10.1109/LRA.2026.3698364

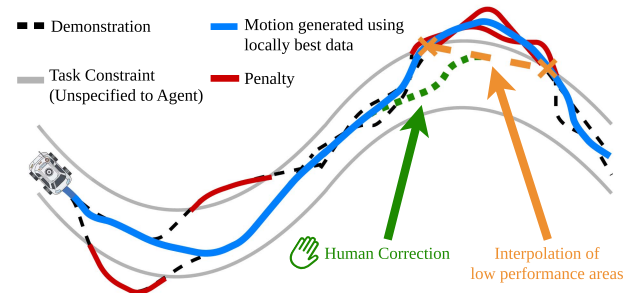


Fig. 1. Conceptual illustration of the proposed LOPAL framework for learning from imperfect demonstrations. The demonstrations (black dashed) contain local imperfections, such as penalties (red) when a vehicle leaves the track (gray). Leveraging the local quality information derived from these penalties allows to generate a motion (blue) that follows the locally best available motion. If all demonstrations are of low quality, the motion is interpolated between adjacent good data points (orange) to estimate a better nominal behavior. During execution, the user is then actively requested to provide better demonstration by correcting this nominal behavior (green).

(LfD) has emerged as a promising alternative, allowing users to intuitively teach new skills to robots through human demonstrations [1]. Despite these advantages, a fundamental challenge in LfD is the strong dependence on the quality of the demonstrations. In this work, we consider demonstration quality in terms of task performance, i.e., the degree to which the demonstrated behavior satisfies the task specifications. If the robot naively replicates the suboptimal demonstrations, it risks reproducing the demonstrated flaws. This challenge is compounded by the inherent variability in human performance during the demonstration. Even a demonstration that is of overall high quality may contain segments of suboptimal behavior [2]. Human task performance is contingent not only upon skill level but also upon extraneous factors such as fatigue [3]. For tasks involving extended operational periods, it is common to observe a decrement in individual performance as the task progresses. Active Learning (AL) allows to reduce the demonstration burden on the user by enabling the robot to actively collect the additional data that it needs to improve its performance. This research endeavors to advance active learning from demonstration by explicitly accounting for local performance, i.e., local variations of task performance in human demonstrations.

A. Related Works

LfD research has begun to recognize the imperfect nature of human demonstrations, yet intra-demonstration variability of quality is still largely overlooked. Early work introduced the concept of learning from failed demonstrations [4], which was later extended to incorporate both successful and unsuccessful

examples [5]. Such a binary classification provides a clear signal for a learning algorithm to differentiate between desired and undesired behaviors. Other methods use continuous score values to rate demonstration quality [6], [7], enabling a more granular prioritization of higher-quality demonstration data during the learning process. In some works, explicit local quality measures are computed directly from the demonstration data [8], [9]. However, these local measures are then collapsed into global scores used to weight the demonstration data for learning. Other work indirectly estimates demonstration quality at a local level from global performance assessments [10], which allows robots to selectively learn from local high-quality data. By relying on global quality ratings and local ratings derived from them, current LfD methods obscure local quality variations within demonstrations. When two demonstrations contain low-quality data in different regions, relying solely on their global assessments cannot reveal which local behaviors are superior. Consequently, their capability to identify and integrate high-quality sub-trajectories into a potentially superior overall behavior is limited.

Certain imperfections of demonstrations have also been handled without explicitly assessing the demonstration quality. In particular, missing data and large temporal or spatial variations between demonstrations were addressed by aligning the demonstrations containing these imperfections to a reference demonstration [11], [12].

Besides pure imitation learning, another prominent direction for improving upon imperfect demonstrations is to combine LfD with Reinforcement Learning (RL). In this paradigm, demonstrations are typically used to bootstrap a Reinforcement Learning agent, for instance by pre-training a policy, which is then fine-tuned through exploration to discover behaviors superior to the ones demonstrated [13], [14]. However, this fine-tuning often requires extensive exploration, which can be time-consuming and in real world scenarios potentially unsafe. The exploration can be guided by human critique of generated trajectories to refine a learned reward [15], shifting the human teaching effort from demonstrations to evaluations of the robot behavior.

In addition to active exploration, which can be performed autonomously by the robot, requesting additional user demonstrations is another approach for obtaining additional data in AL. While user-initiated interventions can be used to demonstrate better behavior [16], actively requesting corrections can reduce the user burden by focusing the human input where it is most needed. This active querying of additional user demonstrations within AL frequently involves dynamically distributing the control between the human and a robot agent (SA) [17], [18]. The robot can actively indicate the need for user input, for example by adjusting the speed [19], modulating the compliance, or providing visual feedback [20]. Multiple active imitation learning approaches request demonstrations based on aleatoric and epistemic uncertainty measures [19], [20]. However, these measures reflect ambiguity over which action to take, not necessarily uncertainty about which demonstrated action may lead to the best task outcome. To the best of our knowledge, existing AL methods have not yet incorporated local variations in demonstration quality to guide data acquisition toward regions where it would most enhance the robot's performance. As a result, data may be over-collected in well-performing regions and under-collected in more difficult ones.

B. Contributions

In this work, we propose LOPAL (Local Performance-aware Active Learning), an AL framework illustrated in Fig. 1, which leverages local performance information to guide both learning from demonstration and active data acquisition, with the aim of enhancing task performance and reducing user effort. Unlike existing LfD methods, which do not consider local quality variations, our approach explicitly leverages local performance information to combine the best parts of multiple imperfect demonstrations, creating a behavior that outperforms any single demonstration. Furthermore, data collection is actively targeted at regions where high-quality demonstration data is missing, thereby aiming to reduce the user's teaching effort. Our contributions are:

- 1) A local performance-aware LfD method that aims to generate trajectories replicating the best demonstrated local behavior.
- 2) An active data acquisition approach, combining autonomous performance-guided local motion modulation with human input through a SA mechanism.

II. LEARNING FROM LOCALLY IMPERFECT DEMONSTRATIONS

We propose a local performance-aware LfD approach (blue section in Fig. 2) to learn efficiently from imperfect demonstrations by leveraging local demonstration quality information. We encode not only the demonstrated trajectories but also their locally varying quality into a GMM (Section II-A). This multimodal data encoded in the GMM allows generating trajectories conditioned on the local quality information, such that the generated trajectory systematically follows demonstration samples associated with the highest local performance (Section II-B). By combining local high-quality data from multiple demonstrations, the system facilitates the synthesis of behaviors that surpass the quality of any single demonstration. Furthermore, these models can be trained from a single demonstration, enabling immediate robot contribution in subsequent task iterations of the incremental teaching procedure. See Algorithm 1 for a summary of our local performance-aware LfD approach.

A. Encoding Demonstrations With Locally Varying Quality

First, both the demonstrated motion and its local quality information are encoded in a GMM. Each of the D demonstrations includes N samples of pose $\xi \in \mathbb{R}^6$. The pose $\xi = [\mathbf{x}, \gamma]^T$ is composed of positions $\mathbf{x} \in \mathbb{R}^3$ and orientations in rotation vector representation $\gamma \in \mathbb{R}^3$. For each sample, a locally varying quality measure is determined, either as a binary penalty $p \in \{0, 1\}$ from constraint violations or as a continuous score $p \in [0, 1]$ (e.g., reflecting motion smoothness or energy efficiency [21]). These local quality measures are then converted into a smoothed local performance $l \in [0, 1]$ for every sample $n \in \{1, \dots, N\}$ using a moving average

$$l[n] = 1 - \frac{1}{\min(N, n+q) - \max(1, n-q) + 1} \sum_{i=\max(1, n-q)}^{\min(N, n+q)} p[i]. \quad (1)$$

The half-window size $q \in \mathbb{N}^+$ balances responsiveness and smoothness when selectively following the locally best demonstration during reference trajectory generation. To represent the progression of the motion, the phase $\phi \in [0, 1]$ is defined

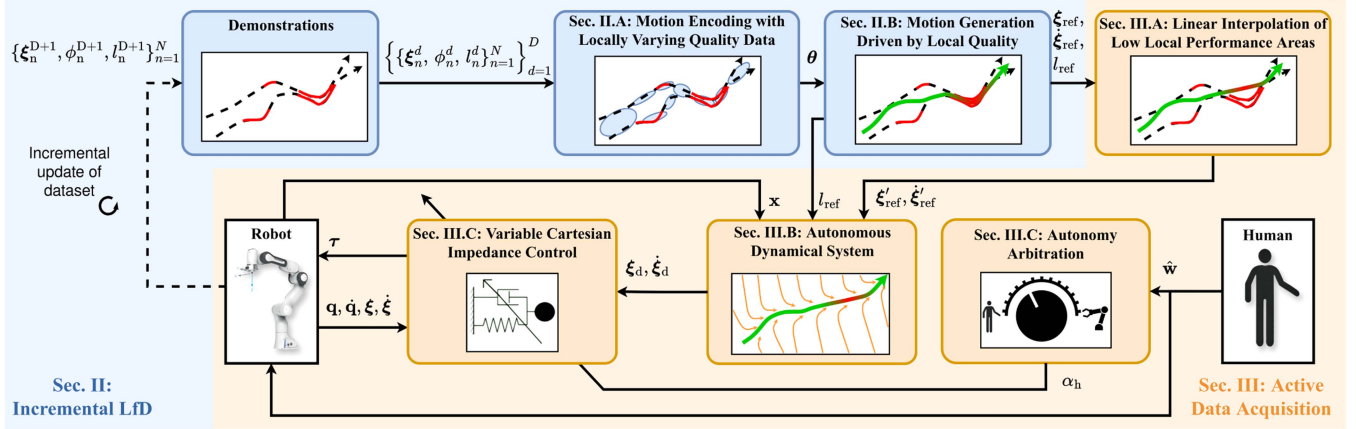


Fig. 2. Overview of proposed LOPAL framework for active learning from imperfect demonstrations using local demonstration quality information. The framework consists of two main components: Incremental LfD from imperfect demonstrations (blue) and active data acquisition (orange). The Incremental LfD component is used to encode an incrementally expanded demonstration dataset and generates trajectories guided by local quality information. The active data acquisition component enables the robot to collect additional data by executing the generated reference motion while the user provides corrections.

based on the normalized cumulative pose distance. Finally, the distribution of the data $\{\{\xi_n^d, \phi_n^d, l_n^d\}_{n=1}^N\}_{d=1}^D$ is approximated by training the GMM in an iterative Expectation-Maximization (EM) procedure. During EM, the log-likelihood of the data is maximized by iteratively updating the GMM parameters

$$\log \mathcal{L}(\theta) = \sum_{d=1}^D \sum_{n=1}^N \log \left(\sum_{k=1}^K \pi_k \mathcal{N}(\xi_n^d, \phi_n^d, l_n^d | \mu_k, \Sigma_k) \right),$$

where $\mathcal{N}$ denotes the normal distribution. The GMM with K number of Gaussian components is described by the parameters $\theta = \{K, \{\pi_k, \mu_k, \Sigma_k\}_{k=1}^K\}$. Here π_k are the priors, μ_k the means, and Σ_k the covariances of the k -th component.

B. Generating Trajectories Driven by Local Performance Information

Using Gaussian Mixture Regression (GMR) [22] we can compute the mean $\bar{\mu}^b(a)$ and variance $\bar{\Sigma}^b(a)$ of the encoded output modalities b given the input a :

$$\begin{aligned} \bar{\mu}^b(a) &= \sum_{k=1}^K h_k(a) \underbrace{(\mu_k^b + \Sigma_k^{ba}(\Sigma_k^a)^{-1}(a - \mu_k^a))}_{\hat{\mu}_k^b(a)}, \\ \bar{\Sigma}^b(a) &= \sum_{k=1}^K h_k(a) \left(\Sigma_k^b - \Sigma_k^{ba}(\Sigma_k^a)^{-1}\Sigma_k^{ab} \right. \\ &\quad \left. + (\hat{\mu}_k^b(a) - \bar{\mu}^b(a)) (\hat{\mu}_k^b(a) - \bar{\mu}^b(a))^T \right). \end{aligned}$$

Here, μ_k^a and μ_k^b are the means of the input and output of the k -th Gaussian component, Σ_k^a and Σ_k^b are the corresponding covariances, and Σ_k^{ba} (and its transpose Σ_k^{ab}) denotes the cross-covariance between output and input modalities. The normalized responsibilities $h_k(a)$ are computed as

$$h_k(a) = \frac{\pi_k \mathcal{N}(a | \mu_k^a, \Sigma_k^a)}{\sum_{i=1}^K \pi_i \mathcal{N}(a | \mu_i^a, \Sigma_i^a)}.$$

Thus, for each of the M reference phase values $\phi_{\text{ref},m}$, $m \in \{1, \dots, M\}$, defined as a sequence linearly increasing from 0 to

Algorithm 1: Local Performance-Driven LfD.

```

1: Input: Demonstration dataset  $\{\{\xi_n^d = [x_n^d, \gamma_n^d], \phi_n^d, p_n^d\}_{n=1}^N\}_{d=1}^D$ 
2:   Parameters:  $K$  (GMM components),  $q$  (half-window size)
3: Output: Reference trajectory  $\{\xi_{\text{ref},m}, l_{\text{ref},m}\}_{m=1}^M$ 
4: ===== ENCODING DEMONSTRATIONS =====
5: for each demonstration  $d \in \{1, \dots, D\}$  do
6:    $\phi^d \leftarrow \text{ComputePhase}(\{\xi_n^d\})$ 
7:    $l^d \leftarrow \text{ComputeLocalPerformance}(\{p_n^d\}, q)$  // eq. (1)
8: end for
9:  $\{\xi_n^d, \phi_n^d, l_n^d\} \leftarrow \text{AlignDemonstrations}(\{\xi_n^d\})$ 
10:  $\{\xi_n^d, \phi_n^d, l_n^d\} \leftarrow \text{NormalizeData}(\{\xi_n^d, \phi_n^d, l_n^d\})$ 
11:  $\theta \leftarrow \text{TrainGMM}(\{\xi_n^d, \phi_n^d, l_n^d\}, K)$ 
12: ===== GENERATING REFERENCE TRAJECTORIES =====
13: for each reference phase  $\phi_{\text{ref},m} \in \text{linspace}(0, 1, M)$  do
14:    $l_{\text{ref},m} \leftarrow \text{ExtractTargetPerformance}(\phi_{\text{ref},m}, \theta)$  // eq. (2)
15:    $\xi_{\text{ref},m} \leftarrow \text{GMR}(\theta, \phi_{\text{ref},m}, l_{\text{ref},m})$  // eq. (3)
16: end for
17: return  $\{\xi_{\text{ref},m}, l_{\text{ref},m}\}_{m=1}^M$ 

```

1, a phase-dependent performance target can be extracted as

$$l_{\text{ref},m} = \bar{\mu}^l(\phi_{\text{ref},m}) + 2\sqrt{\bar{\Sigma}^l(\phi_{\text{ref},m})}. \quad (2)$$

This sets the performance target at two standard deviations above the phase-dependent mean to ensure a high reference performance level; to obtain more conservative performance targets, the mean-offset could be reduced. Using the reference phase and this performance target, we regress the reference pose $\xi_{\text{ref},m} \in \mathbb{R}^6$ as

$$\xi_{\text{ref},m} = \bar{\mu}^\xi([\phi_{\text{ref},m}, l_{\text{ref},m}]^T) \quad (3)$$

By conditioning on the high performance target, the generated movement selectively follows the locally highest-performing demonstrations, thereby preventing degradation through suboptimal demonstrations. Defining the phase via normalized cumulative pose distance and setting ϕ_{ref} to increase linearly ensures that the generated motion maintains nonzero velocity, even when the demonstrations contain pauses.

III. ACTIVE DATA ACQUISITION GUIDED BY LOCAL PERFORMANCE INFORMATION

The active data acquisition component (orange section in Fig. 2) incrementally expands the demonstration dataset, enabling the LfD approach to learn behaviors of higher performance. To request better demonstrations in areas where good data is missing, the robot reduces the speed of the motion execution (Section III-B) and gives a visual indication (e.g. LED light). The user can dynamically take control to make the necessary corrections by applying wrench to the robot, which results in the robot increasing its control compliance (Section III-C). To potentially reduce the required corrections, the reference trajectory is modulated in low-quality areas through interpolation, which leverages neighboring high-quality data (Section III-A). In well-demonstrated areas, the robot can accurately reproduce the learned motion of high local performance. Thus, corrective human input can be focused precisely where the robot has difficulties. See Algorithm 2 for a summary of the active data acquisition approach.

A. Performance-Guided Local Motion Modulation

In low-performance regions, we leverage neighboring high-quality information to create better nominal behavior and reduce the need for user corrections (Fig. 1). If the reference local performance $l_{\text{ref},m}$ of sample $m \in \{1, \dots, M\}$ is below a threshold value $l_{\text{th}} \in [0, 1]$, its position is replaced through linear interpolation between the closest preceding and succeeding high-quality samples:

$$\mathbf{x}'_{\text{ref},m} = \mathbf{x}_{\text{ref},m^-} + \frac{\phi_{\text{ref},m} - \phi_{\text{ref},m^-}}{\phi_{\text{ref},m^+} - \phi_{\text{ref},m^-}} (\mathbf{x}_{\text{ref},m^+} - \mathbf{x}_{\text{ref},m^-}) \quad (4)$$

where $m^- < m$ and $m^+ > m$ denote the indices of the closest samples with $l_{\text{ref},m^\pm} \geq l_{\text{th}}$. This interpolation ensures motions in low local quality areas are modulated using neighboring high-quality samples, while preserving the original behavior in well-demonstrated areas. To ensure a smooth trajectory, a second-order Butterworth low-pass filter with a cutoff frequency of 2 Hz is applied. We compute the reference twist $\dot{\xi}'_{\text{ref},m} = [\dot{\mathbf{x}}'_{\text{ref},m}, \dot{\omega}'_{\text{ref},m}]^T \in \mathbb{R}^6$ by numerically differentiating the regressed reference pose trajectory, where $\omega'_{\text{ref},m} \in \mathbb{R}^3$ is the reference angular velocity.

B. Executing Motions With an Autonomous Dynamical System

To accurately reproduce the reference motion with the robot in a state-dependent manner, an autonomous dynamical system is used. A proportional feedback control term facilitates attraction towards the reference pose trajectory, while a feedforward reference twist drives the system forward

$$\dot{\xi}_d = \mathbf{K}_{DS} \begin{bmatrix} \mathbf{x}'_{\text{ref},j} - \mathbf{x} \\ \mathbf{e}_{\text{rot}}(\gamma'_{\text{ref},j}, \gamma) \end{bmatrix} + s \dot{\xi}'_{\text{ref},j}, \quad (5)$$

where $\mathbf{e}_{\text{rot}}(\gamma_1, \gamma_2) \in \mathbb{R}^3$ is the minimal rotation from γ_2 to γ_1 in rotation vector representation. The streamlines in Fig. 2 illustrate the resulting potential field of the autonomous dynamical system. The attraction stiffness is defined as $\mathbf{K}_{DS} = \text{diag}(k_{DS,t} \mathbf{I}_3, k_{DS,r} \mathbf{I}_3) \in \mathbb{R}^{6 \times 6}$, where $k_{DS,t} \in \mathbb{R}^+$ and $k_{DS,r} \in \mathbb{R}^+$ are the translational and rotational stiffness gains. The reference twist in (5) is scaled by the factor $s \in [s_{\min}, 1]$ to slow down the motion progression when entering

Algorithm 2: Performance-Guided Active Data Acquisition.

```

1: Input: Reference trajectory  $\{\xi_{\text{ref},m}, l_{\text{ref},m}\}_{m=1}^M$ , wrench  $\hat{\mathbf{w}}$ 
2:   Parameters:  $l_{\text{th}}$  (interpolation threshold),  $\mathbf{K}_{DS}$  (stiffness of
   dynamical system),  $s_{\min}$  (min speed),  $\delta$  (lookahead),  $a, b, c$ 
   (autonomy activation parameters),  $r^+, r^-, r_q^-$  (autonomy filter
   parameters),  $\mathbf{K}_{\text{imp}}$  (impedance controller stiffness)
3: Output: New demonstration  $\{\xi_n^{\text{new}}, p_n^{\text{new}}\}_{n=1}^{N^{\text{new}}}$ 
4: ===== INTERPOLATE & FILTER REFERENCE =====
5:  $\{\mathbf{x}'_{\text{ref},m}\} \leftarrow \text{InterpolateLowPerf}(\{\mathbf{x}_{\text{ref},m}, \phi_{\text{ref},m}, l_{\text{ref},m}\}, l_{\text{th}})$ 
   // eq. (4)
6:  $\{\xi'_{\text{ref},m}\} \leftarrow \text{LowPassButterworth}(\{\mathbf{x}'_{\text{ref},m}, \gamma_{\text{ref},m}\}^T, \text{cutoff}$ 
    $= 2 \text{ Hz})$ 
7:  $\{\dot{\xi}'_{\text{ref},m}\} \leftarrow \text{NumericallyDifferentiate}(\{\xi'_{\text{ref},m}\})$ 
8: Initialize:  $j \leftarrow 1, \alpha_h[0] \leftarrow 0, \xi_{\text{new}} \leftarrow \{\}, p_{\text{new}} \leftarrow \{\}$ 
9: while  $j < M$  do
10:  ===== AUTONOMOUS MOTION TRACKING =====
11:   $j \leftarrow \text{FindClosestReference}(\xi, \{\xi'_{\text{ref},m}\})$ 
12:   $s \leftarrow \text{ComputeSpeedScale}(\{l_{\text{ref}}\}, j, \delta, s_{\min})$  // eq. (6)
13:   $\dot{\xi}_d \leftarrow \mathbf{K}_{DS}(\xi'_{\text{ref},j} - \xi) + s \dot{\xi}'_{\text{ref},j}$  // eq. (5)
14:  ===== AUTONOMY ARBITRATION =====
15:   $\alpha_w \leftarrow \text{ComputeWrenchAutonomy}(\hat{\mathbf{w}}, a, b, c)$  // eq. (7)
16:   $\alpha_h \leftarrow \text{FilterAutonomy}(\alpha_h, \alpha_w, r^+, r^-, r_q^-)$  // eq. (8)
17:   $\mathbf{K} \leftarrow (1 - \alpha_h) \mathbf{K}_{\text{imp}}, \mathbf{D} \leftarrow 2\sqrt{\mathbf{K}}$  // Adjust control compliance
18:  ===== DATA ACQUISITION =====
19:  ExecuteControl( $\dot{\xi}_d, \mathbf{K}, \mathbf{D}$ )
20:   $p \leftarrow \text{EvaluatePenalty}$ 
21:   $\xi_{\text{new}} \leftarrow \xi_{\text{new}} \cup \{\xi\}$  // Record pose
22:   $p_{\text{new}} \leftarrow p_{\text{new}} \cup \{p\}$  // Record penalty
23: end while
24: return  $\{\xi_n^{\text{new}}, p_n^{\text{new}}\}_{n=1}^{N^{\text{new}}}$  // Expand dataset

```

regions of low local performance:

$$s = s_{\min} + (1 - s_{\min}) \left(\min_{j' \in \{j, \dots, \min(j+\delta, M)\}} (l_{\text{ref},j'}) \right). \quad (6)$$

Here, $\delta \in \mathbb{N}^+$ is the lookahead window size allowing to proactively slow down before entering the low-performance region, and $s_{\min} \in [0, 1]$ is the minimum velocity scale-down factor. The reduced velocity indicates to the user that the robot is struggling to execute this part of the motion well and gives the human more reaction time to correct the robot's motion. Further, the s value can be used to give additional feedback about low-performance areas to the user, e.g. in the form of visual cues.

The index $j \in \{1, \dots, M\}$ identifies the reference pose $\xi'_{\text{ref}} \in \mathbb{R}^6$ closest to the current pose $\xi \in \mathbb{R}^6$ to couple motion regression with the pose alignment. This ensures local attraction to the reference motion and makes the dynamical system time-invariant. The k-d tree algorithm is used to efficiently compute this pose distance.

C. User Corrections Through Autonomy Arbitration

A shared autonomy approach can allow the human to selectively take over autonomy during incremental teaching in order to correct the robot's task execution. To facilitate human intervention in our LOPAL framework, the robot dynamically adjusts its control compliance to respond to human-applied forces and torques (see Fig. 2). This autonomy arbitration through compliance modulation is controlled by the human autonomy $\alpha_h \in (0, 1)$, which is calculated by filtering the current wrench-based autonomy value $\alpha_w \in (0, 1)$ to enable smooth

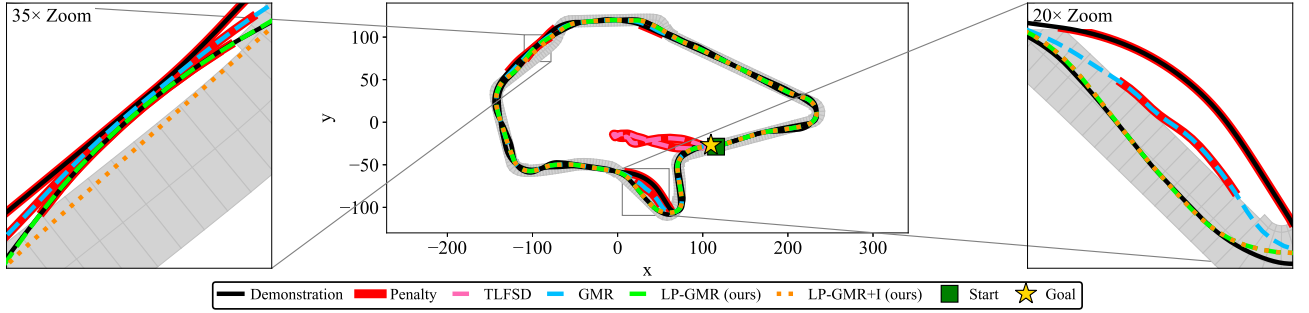


Fig. 3. Illustration of the car driving simulation with two exemplary demonstrations (black). The penalties (red) occurred when the car left the track. The averaging behavior of the GMR (blue dashed), which purely encodes the demonstrated motion, could cause the generated movement to result in penalties, even though one of the demonstrations had shown how to avoid these penalties. Due to these local failures, the motion generated with TLFSD (pink) maximizes distance to the imperfect demonstrations, which often results in even more penalties than in the demonstrated motions. In the proposed local performance-aware approach (green dashed), the information about the location of penalties was effectively used to avoid these penalty areas. Experiment results demonstrated that interpolating regions lacking high quality demonstration data (orange dotted) allowed to further reduce penalties.

and adaptable autonomy arbitration. First, we compute the value α_w using the wrench $\hat{w} \in \mathbb{R}^6$ [18]:

$$\alpha_w = 0.5 \left(1 + \tanh \left(\frac{3}{ca} \left(\|\hat{w}\| - \frac{a}{2} - b \right) \right) \right). \quad (7)$$

Here, a controls how quickly the activation increases with $\|\hat{w}\|$, and b sets the deadzone below which the response is suppressed. To enforce the tolerance $\alpha_w(\|\hat{w}\| = b) = 1 - \alpha_w(\|\hat{w}\| = a + b) = \epsilon$, we choose $c = \frac{3}{2 \operatorname{atanh}(1-2\epsilon)}$. The wrench $\hat{w}$ is obtained from a force-torque sensor mounted at the robot wrist, which enables high sensitivity to human-applied interaction wrench.

To avoid discontinuities, a recursive filter is used to compute α_h at the i -th control loop step:

$$\alpha_h[i] = \alpha_h[i-1] + r_\alpha[i](\alpha_w[i] - \alpha_h[i-1]). \quad (8)$$

where the filter gain $r_\alpha \in \mathbb{R}^+$ is defined as

$$r_\alpha[i] = \begin{cases} r^+, & \text{if } \alpha_w[i] > \alpha_w[i-1], \\ r^- + r_q^-(1 - \alpha_w[i])^2, & \text{otherwise.} \end{cases} \quad (9)$$

To achieve a rapid human autonomy increase when force is applied and a slower decrease when force is released, we set $r^+ > r^-$. The quadratic term in (9) with gain r_q^- facilitates a fast reduction for small α_w values, which can result from briefly applied low forces that may be transient and less indicative of a persistent human intention to take control. Finally, α_h is used to modulate the stiffness $\mathbf{K} = (1 - \alpha_h)\mathbf{K}_{\text{imp}} = (1 - \alpha_h) \operatorname{diag}(k_t \mathbf{I}_3, k_r \mathbf{I}_3) \in \mathbb{R}^{6 \times 6}$ and damping $\mathbf{D} = 2\sqrt{\mathbf{K}} \in \mathbb{R}^{6 \times 6}$ of a Cartesian variable impedance controller [18], [23]. An energy-tank-based approach [18] was used to ensure the passivity of the controller and thus increase safety when operating in close proximity to humans [24].

IV. SIMULATION EVALUATION

We evaluated our local performance-aware LfD method on a car driving task using human demonstrations on a 2D track. Specifically, we investigated the ability to generate trajectories that avoid local penalties while learning efficiently from an incrementally expanding set of demonstrations. In this work, all statistical analyses were performed using Wilcoxon signed-rank tests, as the data were paired and not normally distributed. Effect sizes are reported as rank-biserial correlation (r), where 0 indicates no difference and ± 1 indicates one condition consistently ranked higher than the other, with 95% bootstrap confidence intervals (CI).

A. Simulation Setup

The task required driving on a 2D track while avoiding penalties, which were incurred whenever the vehicle's center left the road. The demonstrations were collected by controlling the car via keyboard on a fixed track in the OpenAI Gym car racing environment. The training data comprised 15 demonstrations, each containing a 2D position trajectory with corresponding phase and local performance values. In 50 experimental sessions, each time we started with one of the demonstrations and gradually added more demonstrations one-by-one to the training dataset in random order. This allowed us to assess the ability to learn efficiently from an incrementally expanded set of imperfect demonstrations.

B. Compared LfD Methods

1) *Reinforcement Learning From Prior Data (RLPD)* [14]: Learning combines LfD and RL by sampling equally from a demonstration buffer and an online replay buffer during training. Local performance information is used as a dense reward.

2) *Trajectory Learning From Failed and Successful Demonstrations (TLFSD)* [5]: Demonstration quality is assessed globally by classifying each demonstration as successful or unsuccessful. Separate GMMs are trained for each set, and trajectories are generated by minimizing deviation from the successful GMM while maximizing deviation from the failed GMM using Mahalanobis distance.

3) *Vanilla GMM/GMR (GMR)*: The GMM is trained on motion data, while the performance information is excluded from the training data. Consequently, motion generation is performed using only the phase variable as regression input.

4) *Local Performance Pruned GMM/GMR (P-GMR)*: Local performance information is used to prune the penalized samples from the training data such that the GMM is trained only on high-quality motion data. Motion generation uses the phase variable as regression input, as in the vanilla GMR.

5) *Local Performance-Aware GMM/GMR (LP-GMR)*: Local performance information is encoded in the GMM together with the motion data. This allows to condition the motion generation both on phase and on high local performance such that the reference motion follows the locally best available motion (Section II-B).

6) *Local Performance-Aware GMM/GMR With Interpolation (LP-GMR+I)*: Extending LP-GMR, the encoded local

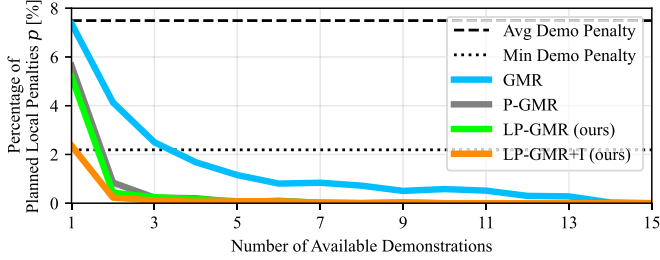


Fig. 4. Reduction of penalties for generated motions as an increasing number of demonstrations become available (fewer penalties are better). Both ablations using local performance information (LP-GMR green and P-GMR gray) achieved better performance using fewer demonstrations compared to the base GMR approach (blue) which was not informed by demonstration quality. Additional performance benefits were achieved by interpolating regions lacking high-quality demonstration data (LP-GMR+I orange). As a reference, the average (black dashed) and minimum (black dotted) penalties from demonstrations are shown.

performance information is used not only for motion regression but also to interpolate motions in regions lacking high-quality demonstration data (Section III-A).

C. Simulation Results

If at least one successful (i.e. penalty-free) demonstration exists, TLFSD can replicate the motion without penalties. However, for challenging tasks, especially with long horizons, it can be difficult to provide a full demonstration without a single penalty, which would result in the demonstration being labeled as failed. If only failed demonstrations are available, TLFSD often generates trajectories with many penalties, since it maximizes the distance from these imperfect demonstrations (pink trajectory in Fig. 3). While iterative generative learning can improve results, it requires careful parameter tuning and does not guarantee a successful motion within a fixed number of iterations, particularly for long or complex motions. All 15 demonstrations exhibited imperfections on the full track. Since TLFSD generated even more penalties than these demonstrations, it was not considered in the detailed evaluation on the full track (Fig. 4). Unlike TLFSD, which relies on global assessment of demonstration success, our approach leverages local quality cues to handle partial imperfections. The LfD approaches leveraging local performance information can generate motions that selectively follow the locally best available demonstrated behavior (see right zoomed view of Fig. 3). Conversely, base GMR, which does not leverage any performance information, is prone to replicate low-performance behaviors present in most demonstrations due to its averaging nature. If all demonstrations show penalties in a region, interpolating between neighboring high-quality samples can reduce or even avoid these penalties, as illustrated in the left zoomed view of Fig. 3. The LfD methods leveraging local performance information (P-GMR and LP-GMR) yielded significantly fewer penalties than the base GMR ($p < 0.001$, smallest $|r| = 0.98$, $CI_{95\%} = [0.95, 0.99]$). This can be attributed to their capability to utilize complementary high-quality local demonstration data. As mentioned in Section II-A, local performance can be derived not only from binary penalties, as in this task, but also from continuous ratings. In contrast to LP-GMR, when applying P-GMR to such continuous ratings, there is no principled criterion for choosing the pruning threshold, and the retained samples contribute equally to motion generation regardless of their differing quality levels. Compared to the

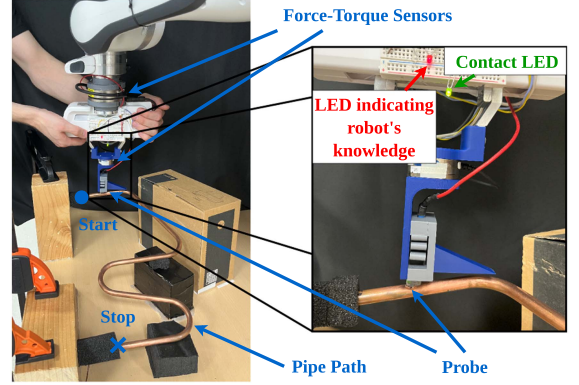


Fig. 5. Pipe inspection experiment setup used for the validation of LOPAL. Here, participants were tasked to trace the pipe path with the probe from start to end, while maintaining contact between probe and pipe which is indicated by the green LED. In one of the teaching modes, a red LED indicates the robot's limited knowledge of high-quality data to the user. Two force-torque sensors are used to determine the forces applied by the user.

performance leveraging ablations, linear interpolation of low-performance regions (LP-GMR+I) further improved learning efficiency when using only one or two demonstrations, resulting in an additional penalty reduction ($p < 0.006$, smallest $|r| = 0.73$, $CI_{95\%} = [0.38, 0.97]$). RLPD achieved penalty-free task performance but required 430 ± 140 (mean $\pm$ std) additional exploration episodes ($129,013 \pm 46,852$ steps); such extensive exploration may not pose a problem in simulation but could be impractical or potentially unsafe in real-world deployment.

These results underscore the capability of the proposed LfD approach to enhance learning efficiency from imperfect demonstrations by using local performance information.

V. REAL-WORLD EXPERIMENT

We evaluated our proposed LOPAL approach in a real-world incremental teaching scenario on a Franka Research 3 robot. Specifically, we investigated how teaching with the proposed method affects the burden on the user. In addition, we evaluated our proposed LfD approach in this real-world experiment, similar to the evaluation performed for the simulation. The experiment was conducted with 8 participants (no prior power analysis), who confirmed their participation in the experiment by signing a written informed consent form. The ethics proposal for this study was approved by the TU Wien Research Ethics Committee (case number 045_19012024_TUWREC).

A. Experiment Setup

In this real-world experiment, users had to kinesthetically teach the robot a pipe inspection task. Participants were asked to trace a copper pipe while maintaining contact between the end-effector-mounted probe and the pipe, see Fig. 5. The loss of contact between pipe and probe was considered a penalty. Maintaining continued contact is important because sensors used for inspection often require direct contact or a fixed offset for proper functioning. Because the pipe was mounted in an inclined orientation and supporting structures had to be avoided, the position and orientation of the end effector needed to be controlled in order to trace the pipe path without losing contact. For every teaching mode, an initial demonstration with low performance was provided such that in the shared autonomy approaches the robot could contribute from the first trial and

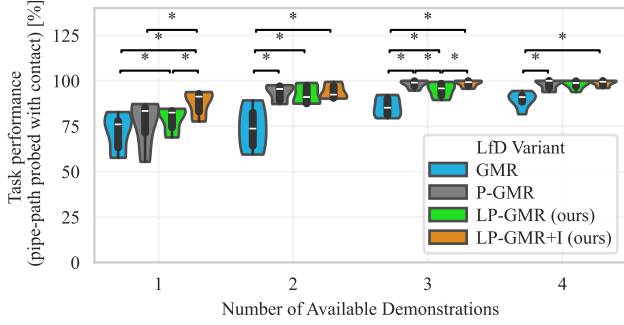


Fig. 6. Comparison of LfD methods in terms of task performance as the number of demonstrations increased incrementally. Leveraging local performance information (P-GMR gray, LP-GMR green, LP-GMR+I orange) allowed to achieve higher task performance with fewer demonstrations compared to the GMR baseline (blue). When only a single demonstration was available, interpolation of low-quality regions (LP-GMR+I) provided further performance benefits.

TABLE I

THE EMPIRICALLY SELECTED PARAMETERS USED FOR THE PROPOSED LOPAL FRAMEWORK IN THE REAL-WORLD EXPERIMENT

$r^+ = 0.03$	$k_{DS,t} = 47 \text{ s}^{-1}$	$k_t = 2500 \text{ N m}^{-1}$	$a = 10$
$r^- = 0.002$	$k_{DS,r} = 27 \text{ s}^{-1}$	$k_r = 15 \text{ N m rad}^{-1}$	$b = 3$
$r_q = 0.02$	$K = 95$	$f_c = 150 \text{ Hz}$	$c = 0.434$
$s_{\min} = 0.4$	$\delta = 45$	$q = 3$	$l_{th} = 0.7$

the starting behavior was the same for all participants, making results more comparable. In all teaching modes, three demonstrations were collected sequentially, incrementally expanding the dataset and updating the model after each trial. After collecting the demonstrations with each teaching mode, participants assessed their perceived mental and physical workload using the NASA-TLX questionnaire. To compare different LfD methods (Section IV-B) regarding their efficiency to learn from imperfect demonstrations, the robot autonomously executed the generated behaviors without user corrections while the demonstration set was incrementally expanded (evaluated on six datasets of four demonstrations). Performance in this task was evaluated in terms of the percentage of pipe path traced without penalty. The forces exerted by the user were determined by comparing the sensor values of the two force-torque sensors shown in Fig. 5. The parameters used for our LOPAL approach in this experiment are summarized in Table I.

B. Compared Kinesthetic Teaching Methods

1) *Gravity Compensation (GC)*: The robot remains passive such that the human is in full control throughout the task execution.

2) *Shared Autonomy With GMR (SA + GMR)*: The robot executes the behavior which was learned without leveraging performance information (GMR). By applying wrench to the robot, the user can dynamically take control to correct the robot's task execution (Section III-C).

3) *Shared Autonomy With P-GMR (SA + P-GMR)*: The robot executes the motion learned from pruned, high-quality data (P-GMR). Our proposed SA method allows the user to correct the robot's task execution.

4) *LOPAL (ours)*: Our wrench-based shared autonomy method is combined with the proposed LfD method (LP-GMR+I). During teaching, active performance feedback is

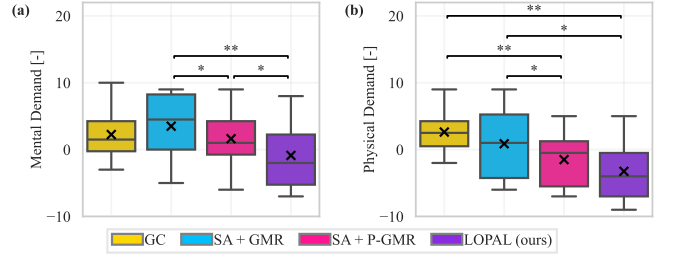


Fig. 7. Subjective ratings of (a) mental and (b) physical demand for the different teaching variants: GC (yellow), SA + GMR (blue), SA + P-GMR (pink), and LOPAL (purple).

provided through speed modulation and visual indication (red LED, Fig. 5) to prompt user corrections.

C. Experiment Results

1) *Comparison of LfD Approaches*: Fig. 6 illustrates how task performance evolves as the number of available demonstrations increases. In this real-world experiment, we compared the LfD methods GMR, P-GMR, LP-GMR, and our LP-GMR+I (described in Section IV-B).

Consistent with the simulation results, the real-world experiment results also showed that all LfD approaches leveraging local performance information tended to achieve better performance than baseline GMR. Our proposed approach (LP-GMR+I) consistently achieved better performance than base GMR ($p = 0.031$, $|r| = 1$, $CI_{95\%} = [1.00, 1.00]$), achieving an average performance improvement of up to 27.31% depending on the number of demonstrations. When only a single demonstration was available, LP-GMR+I's interpolation mechanism reduced the low-performance areas further compared to P-GMR and LP-GMR ($p = 0.031$, smallest $|r| = 1$, $CI_{95\%} = [1.00, 1.00]$).

2) *Comparison of Teaching Approaches*: As illustrated in Fig. 7(b), participants reported reduced physical task load with the teaching approaches that leverage local performance information (SA + P-GMR and LOPAL) compared to baselines (GC and SA + GMR) which do not use performance information ($p < 0.028$, smallest $|r| = 0.83$, $CI_{95\%} = [0.21, 0.97]$). These subjective results are also reflected quantitatively, as teaching with GC and SA + GMR resulted in higher energy injected by the user ($p < 0.023$, smallest $|r| = 0.89$, $CI_{95\%} = [0.50, 1.00]$), see Fig. 8. The high physical demand with these two approaches may be due to the robot requiring much user input, either to demonstrate the full task (GC) or to make many corrections due to inefficient learning (SA + GMR). Furthermore, in the first teaching trial, which required the most correction, the maximum force applied by users was lower with LOPAL than with SA + P-GMR ($p = 0.039$, $|r| = 0.83$, $CI_{95\%} = [0.17, 1.00]$). Since LOPAL slows down in areas with low local demonstration quality, users may no longer need to apply large initial forces when making corrections to manually slow the robot down. This could explain the tendency of LOPAL's subjective physical demand being rated lower, seen in Fig. 7(b). Additionally, participants reported lower mental demand with LOPAL than with the other SA-based teaching variants (SA + GMR and SA + P-GMR) ($p < 0.027$, smallest $|r| = 0.90$, $CI_{95\%} = [0.35, 0.99]$). This may be attributed to the active user feedback, which likely helped users to focus their corrections where they were most

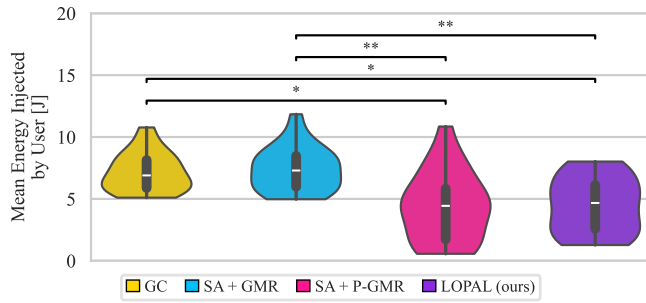


Fig. 8. Mean energy injected by the participants during the teaching trials when using different teaching variants. When incrementally teaching with approaches that leverage local performance information (SA + P-GMR pink and LOPAL purple), users expended less energy than with GC (yellow) and SA + GMR (blue), which do not use local performance information.

needed. Verbal reports from participants provided insight on why the mental demand with GC resulted in similar ratings as with the SA-based teaching approaches: while shared autonomy reduced the mental effort associated with directly performing the task, it simultaneously introduced a new form of cognitive demand — monitoring and reacting to the robot’s autonomous behavior.

3) **Key Findings:** Our experiment evaluation shows that in a real-world robotic task, the use of local quality information in LfD can enhance learning effectiveness, which can be further enhanced by bridging regions of low local performance through interpolation.

This efficient learning approach, which leverages local performance information, enabled a significant reduction in the physical teaching effort. Furthermore, participant ratings indicate that providing feedback on missing high-quality data through speed reduction and visual cues has the potential to decrease the user’s mental load during incremental teaching. Overall, these results suggest that our proposed LOPAL framework can increase learning efficiency from imperfect demonstrations while simultaneously reducing the teaching burden on the user.

VI. CONCLUSION

This letter presents LOPAL – a novel local performance-aware active learning from demonstration framework with two main contributions: (1) A LfD approach that leverages locally varying performance information to generate a reference behavior that selectively replicates the best demonstrated local behaviors, thereby allowing it to outperform the original demonstrations. (2) An active data collection method that facilitates incremental learning by prompting the user to correct the execution of this reference behavior in order to acquire additional informative samples. Simulation and experiment results highlight the benefit of our proposed framework. LOPAL not only improved task performance but also reduced the demonstration effort for the user. Future work could investigate the use of other quality measures for applying LOPAL to a wider range of tasks or extending the approach to larger state spaces, like dual-arm manipulation tasks.

ACKNOWLEDGMENT

The authors acknowledge TU Wien Bibliothek for financial support through its Open Access Funding Program.

REFERENCES

- [1] D. Lee and C. Ott, “Incremental kinesthetic teaching of motion primitives using the motion refinement tube,” *Auton. Robots*, vol. 31, pp. 115–131, Jul. 2011.
- [2] Y. Yu, H. Xu, Y. Xu, F. Lu, and M. Li, “Increased intra-individual variability as a marker of executive dysfunction in generalized anxiety disorder,” *Front. Psychiatry*, vol. 13, Mar. 2022, Art. no. 532778.
- [3] Y. T. Sam, E. Hedlund-Botti, M. Natarajan, J. Heard, and M. Gombolay, “The impact of stress and workload on human performance in robot teleoperation tasks,” *IEEE Trans. Robot.*, vol. 40, pp. 4725–4744, 2024.
- [4] D. H. Grollman and A. Billard, “Donut as I do: Learning from failed demonstrations,” in *Proc. IEEE Int. Conf. Robot. Automat.*, 2011, pp. 3804–3809.
- [5] B. Hertel and S. R. Ahmadzadeh, “Learning from successful and failed demonstrations via optimization,” in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2021, pp. 7807–7812.
- [6] Z. Cao and D. Sadigh, “Learning from imperfect demonstrations from agents with varying dynamics,” *IEEE Robot. Automat. Lett.*, vol. 6, no. 3, pp. 5231–5238, Jul. 2021.
- [7] H. Hoang, T. Mai, and P. Varakantham, “SPRINQL: Sub-optimal demonstrations driven offline imitation learning,” in *Proc. Adv. Neural Inf. Process. Syst.*, 2024, pp. 136837–136872.
- [8] M. Bilal, N. Lipovetzky, D. Oetomo, and W. Johal, “Beyond success: Quantifying demonstration quality in learning from demonstration,” in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2024, pp. 5120–5127.
- [9] M. Sakr, J. Zhang, H. F. M. V. d. Loos, D. Kulić, and E. Croft, “Consistency matters: Defining demonstration data quality metrics in robot learning from demonstration,” *J. Hum.-Robot Interaction*, vol. 15, pp. 1–31, Dec. 2025.
- [10] Z.-H. Yin and P. Abbeel, “Offline imitation learning through graph search and retrieval,” in *Proc. Robot., Sci. Syst. XX*, 2024.
- [11] A. Pervez, A. Ali, J.-H. Ryu, and D. Lee, “Novel learning from demonstration approach for repetitive teleoperation tasks,” in *Proc. IEEE World Haptics Conf.*, 2017, pp. 60–65.
- [12] A. Pervez, H. Latifee, J.-H. Ryu, and D. Lee, “Motion encoding with asynchronous trajectories of repetitive teleoperation tasks and its extension to human-agent shared teleoperation,” *Auton. Robots*, vol. 43, pp. 2055–2069, Dec. 2019.
- [13] A. Ren et al., “Diffusion policy policy optimization,” in *Proc. Int. Conf. Learn. Representations*, 2025, pp. 77288–77329.
- [14] P. J. Ball, L. Smith, I. Kostrikov, and S. Levine, “Efficient online reinforcement learning with offline data,” in *Proc. 40th Int. Conf. Mach. Learn.*, 2023, pp. 1577–1594.
- [15] E. Hirota et al., “Active reward learning and iterative trajectory improvement from comparative language feedback,” *Int. J. Robot. Res.*, Nov. 2025, Art. no. 02783649251394742.
- [16] J. Luo, C. Xu, J. Wu, and S. Levine, “Precise and dexterous robotic manipulation via human-in-the-loop reinforcement learning,” *Sci. Robot.*, vol. 10, Aug. 2025, Art. no. eads5033.
- [17] Y. Michel, Z. Li, and D. Lee, “A learning-based shared control approach for contact tasks,” *IEEE Robot. Automat. Lett.*, vol. 8, no. 12, pp. 8002–8009, Dec. 2023.
- [18] S. Jadav, J. Heidersberger, C. Ott, and D. Lee, “Shared autonomy via variable impedance control and virtual potential fields for encoding human demonstrations,” in *Proc. IEEE Int. Conf. Robot. Automat.*, 2024, pp. 15151–15157.
- [19] T. Eiband, C. Willibald, I. Tannert, B. Weber, and D. Lee, “Collaborative programming of robotic task decisions and recovery behaviors,” *Auton. Robots*, vol. 47, pp. 229–247, Feb. 2023.
- [20] C. Celemin and J. Kober, “Knowledge- and ambiguity-aware robot learning from corrective and evaluative feedback,” *Neural Comput. Appl.*, vol. 35, pp. 16821–16839, Aug. 2023.
- [21] S. Zhang, Q. Xia, M. Chen, and S. Cheng, “Multi-objective optimal trajectory planning for robotic arms using deep reinforcement learning,” *Sensors*, vol. 23, Jun. 2023, Art. no. 5974.
- [22] S. Calinon and D. Lee, “Learning Control,” in *Humanoid Robotics: A Reference*, P. Vadakkepat and A. Goswami Eds., Dordrecht, Netherlands: Springer, 2019, pp. 1261–1312.
- [23] H. Xue, Y. Michel, and D. Lee, “A shared control approach based on first-order dynamical systems and closed-loop variable stiffness control,” *J. Intell. Robotic Syst.*, vol. 109, Dec. 2023, Art. no. 85.
- [24] Y. Michel, M. Saveriano, and D. Lee, “A novel safety-aware energy tank formulation based on control barrier functions,” *IEEE Robot. Automat. Lett.*, vol. 9, no. 6, pp. 5206–5213, Jun. 2024.