Check for updates

Article

The International Journal of Robotics Research
2021, Vol. 40(12-14) 1331–1351
© The Author(s) 2021
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/02783649211021869
journals.sagepub.com/home/ijr
SAGE

Task space adaptation via the learning of gait controllers of magnetic soft millirobots

Sinan O. Demir[1,2*], Utku Culha[1*], Alp C. Karacakol[1,3], Abdon Pena-Francesch[1,4], Sebastian Trimpe[5,6] and Metin Sitti[1]

Abstract

Untethered small-scale soft robots have promising applications in minimally invasive surgery, targeted drug delivery, and bioengineering applications as they can directly and non-invasively access confined and hard-to-reach spaces in the human body. For such potential biomedical applications, the adaptivity of the robot control is essential to ensure the continuity of the operations, as task environment conditions show dynamic variations that can alter the robot's motion and task performance. The applicability of the conventional modeling and control methods is further limited for soft robots at the small-scale owing to their kinematics with virtually infinite degrees of freedom, inherent stochastic variability during fabrication, and changing dynamics during real-world interactions. To address the controller adaptation challenge to dynamically changing task environments, we propose using a probabilistic learning approach for a millimeter-scale magnetic walking soft robot using Bayesian optimization (BO) and Gaussian processes (GPs). Our approach provides a data-efficient learning scheme by finding the gait controller parameters while optimizing the stride length of the walking soft millirobot using a small number of physical experiments. To demonstrate the controller adaptation, we test the walking gait of the robot in task environments with different surface adhesion and roughness, and medium viscosity, which aims to represent the possible conditions for future robotic tasks inside the human body. We further utilize the transfer of the learned GP parameters among different task spaces and robots and compare their efficacy on the improvement of data-efficient controller learning.

Keywords

Soft robotics, adaptive locomotion, controller learning, Bayesian optimization, Gaussian processes, transfer learning

1. Introduction

Soft robots are composed of highly deformable soft materials exhibiting programmable shape change, mechanical compliance, and high degrees of freedom, which are hard to achieve using rigid materials (Majidi, 2014). The easier access to novel fabrication methods further allows the engineering of stimuli-responsive soft materials that enable new functionalities for soft robots in multiple length scales (Shen et al., 2020). Biology remains to be a source of inspiration for the design, control, and behavior of soft robots (Laschi et al., 2016) and provides templates for new application areas in multi-terrain locomotion (Calisti et al., 2017), adaptive manipulation (Hughes et al., 2016), sensing (Iida and Nurzaman, 2016), human-assistive wearable systems (Walsh, 2018), and biomedicine (Cianchetti et al., 2018). Soft robots also enable safe human–robot physical interaction due to their physical compliance and the mechanical dampening of excess forces (Polygerinos et al., 2017), which otherwise require additional computational effort in conventional robotic systems (Haddadin

[1]Physical Intelligence Department, Max Planck Institute for Intelligent Systems, Stuttgart, Germany
[2]Stuttgart Center for Simulation Science (SC SimTech), University of Stuttgart, Stuttgart, Germany
[3]Department of Mechanical Engineering, Carnegie Mellon University, Pittsburgh, PA, USA
[4]Department of Materials Science and Engineering, Robotics Institute, University of Michigan, Ann Arbor, MI, USA
[5]Institute for Data Science in Mechanical Engineering, RWTH Aachen University, Aachen, Germany
[6]Intelligent Control Systems Group, Max Planck Institute for Intelligent Systems, Stuttgart, Germany
[*]These authors have contributed equally.

Corresponding author:
Metin Sitti, Physical Intelligence Department, Max Planck Institute for Intelligent Systems, 70569 Stuttgart, Germany.
Email: sitti@is.mpg.de

et al., 2017). Small-scale (i.e., $\leqslant 1$ cm) untethered soft robots have further potential application areas in medicine owing to their ability to access enclosed small spaces non-invasively (Sitti, 2018) and the embodiment of functionalized materials enabling targeted drug delivery, diagnostics, and surgery (Cianchetti et al., 2018).

Despite their exciting potential and new capabilities, soft robots face challenges that arise from the nature of their soft materials, such as having virtually infinite degrees of freedom, being prone to fabrication-dependent performance variabilities that cause more significant effects at the smaller scale, and nonlinear material behavior (e.g., hysteresis, creep). Moreover, physical interactions of soft-bodied robots with their operation environment, such as solid or fluid operation medium, are very hard to model due to complex fluid–structure interactions, soft body dynamics, and contact mechanics. The combination of these aspects renders the application of conventional modeling and control methods challenging for soft robots, especially for untethered systems at the small scales (Rus and Tolley, 2015). One of the most widely used methods is the employment of the constant curvature (CC) models that utilize the well-established beam theories to model the kinematics and dynamics of axisymmetrically bending soft robotic systems (Della Santina et al., 2020; Webster and Jones, 2010). Alternatively, analytical approaches using Cosserat rod models (Renda et al., 2018) and geometrically exact models have been suggested for continuum robots (Grazioso et al., 2019). Simulation techniques build upon these modeling methods as in the finite-element methods (FEMs), which construct continuum robot structures using a chain of rigid elements connected with tunable spring–damper mechanisms (Chenevier et al., 2018; Goury and Duriez, 2018). Numerical approaches using voxel-based representations (Hiller and Lipson, 2014) and discrete differential geometries (DDGs) (Huang et al., 2020) improve the computation time of soft robotic simulations at the expense of nonlinear dynamics precision. These models and simulation tools typically allow the implementation of static and dynamic controllers for continuum robots on a larger scale (Thuruthel et al., 2018). However, the physical application of these closed-loop controllers depends on the continuous sensing of body deformations from embedded sensors and highly responsive actuators, and computationally heavy model solutions, which are conditions that may not be met for untethered soft robots at the small scales (Rich et al., 2018). Therefore, the soft robotic platforms that successfully employ the analytical models at the small scales still depend on either open-loop (Gu et al., 2020; Lu et al., 2018; Ren et al., 2019; Wu et al., 2019) or manually applied (Kim et al., 2019) controllers. In particular, for those robots targeting medical applications, their dynamically changing and deformable task environments, fabrication-based variations, and material degradation over prolonged use significantly alter their robotic function performances and pose challenges for the conventional control strategies (Sitti, 2018). The combination of these challenges makes the machine learning-based, adaptive, and data-efficient (i.e., using as few experiments as possible) control methods more desirable for untethered small-scale soft robots.

Data-driven machine learning methods may provide alternative solutions for the design and control of soft robots in the lack of existing analytical or numerical models that describe their underlying kinematics, dynamics, and functions (Chin et al., 2020). One common approach is to learn these models by gathering data from robot experiments and training a neural network (NN) architecture (Bern et al., 2020; Hyatt et al., 2019; Thuruthel et al., 2019). However, the need for data efficiency, i.e., the ability to learn from only a few experimental trials, presents a core challenge for such methods (Chatzilygeroudis et al., 2019). Conversely, Bayesian optimization (BO) (Ghahramani, 2015; Shahriari et al., 2015) allows for the maximization of a performance function using a small number of physical experiments. BO typically employs Gaussian processes (GPs) (Rasmussen and Williams, 2006) as a probabilistic model of the latent objective function. Although no explicit dynamics model is needed, GPs allow for incorporating information as probabilistic priors, thus reducing the experimental data requirements. There are emerging examples that demonstrate the application of this approach to optimize the locomotion performance of robots on different length scales (Calandra et al., 2016; Liao et al., 2019; Marco et al., 2020; Yang et al., 2018). Despite its potential, there are only a few examples that apply this method to address the controller challenge for untethered soft robots, such as in the gait exploration of a tensegrity system (Rieffel and Mouret, 2018), and the optimization of an undulating motion of a microrobot (von Rohr et al., 2018).

For cases where the training and testing domains show differences in terms of features or data distribution, transfer learning (TL) methods may provide further improvements in data-efficient learning and adaptation to new test cases (Pan and Yang, 2009). Within the BO applications that employ GPs, the prior knowledge can be transferred as GP priors (Raina et al., 2006) and hyperparameters (Perrone et al., 2019) from the trained domains to provide predictive information about the unknown features and distributions in the new test domains. In robotics, TL is typically employed as the transfer of the models of kinematics and dynamics between simulated and physical platforms of conventional rigid robotic systems, such as manipulators (Devin et al., 2017; Makondo et al., 2018), humanoids (Delhaisse et al., 2017), and quadrotor platforms (Helwa and Schoellig, 2017). However, the application of TL on soft robotics systems is still in its early infancy (Schramm et al., 2020).

In our recent work in Culha et al. (2020), we demonstrated the controller learning of walking soft millirobots using BO and GPs and showed the improvement of learning efficiency in means of transferring prior mean information between robots as a TL application example. We

followed the initial example by von Rohr et al. (2018), which designed a learning scheme by comparing different GP priors and BO settings on generating a semi-synthetic dataset that represents the estimated gait controller space and used this estimation to optimize the one-dimensional crawling gait of a light-driven soft microrobot. In our work in Culha et al. (2020), we adopted the magnetic soft millirobots from Hu et al. (2018) that lacked sufficient predictive kinematic models and was therefore controlled with an open-loop system whose multi-dimensional parameters were tuned manually. We showed that these robots suffered performance inconsistencies due to the fabrication reproducibility issues, material degradation over prolonged experiments, and environmental disturbances, which limited the derivation of a deterministic kinematic model and the application of relevant model-based controllers. Therefore, we applied BO and GPs to directly learn the controller parameters while optimizing the stride length performance of these robots and employed TL methods to improve learning efficiency using a small number of physical experiments.

In this study, we extend our previous work in Culha et al. (2020) and provide an in-depth analysis on using BO and GPs (in particular for TL) to directly and efficiently learn the controller parameters of the magnetic soft millirobots' walking gait on task spaces emulating bio-medical application environments (Figure 1). First, we introduce our new automated and closed-loop experimental platform that can run the robot learning experiments repeatedly and reliably to eliminate the influence of any human intervention, which caused further material degradation and consequent performance inconsistencies in Hu et al. (2018) and Culha et al. (2020). We start with using an exhaustive search on the two-dimensional (2D) gait controller parameter space of the millirobot and generating benchmark datasets that show the stride length performances of three different robots on three different walking surfaces. We use this benchmark data to learn the optimum gait controllers using BO and GPs, and then to compare the influence of four different TL methods on the improvement of learning efficiency. We choose the best performing TL method from these experiments and use it with the BO and GPs to learn the walking gait controller parameters on a wide range of task spaces. We test our robots on task spaces with different surface roughness and friction, and liquid medium viscosity to emulate the conditions inside the human body for future target operations. Our results reveal that the direct controller learning with BO and GPs allows for adaptation to different task spaces for small-scale untethered soft robots that are prone to fabrication-, material-, and interaction-dependent performance variabilities. We also show that the effective use of TL improves this adaptation by exploring a larger set of successful walking gait controllers within a limited number of physical experiments despite the significantly changing task space conditions. The methodology we present in this study can be used for controlling future small-scale soft robot applications for medical operations that require a data-efficient controller learning system and quick adaptation to the changing task environments. In summary, the main contributions of our work are:

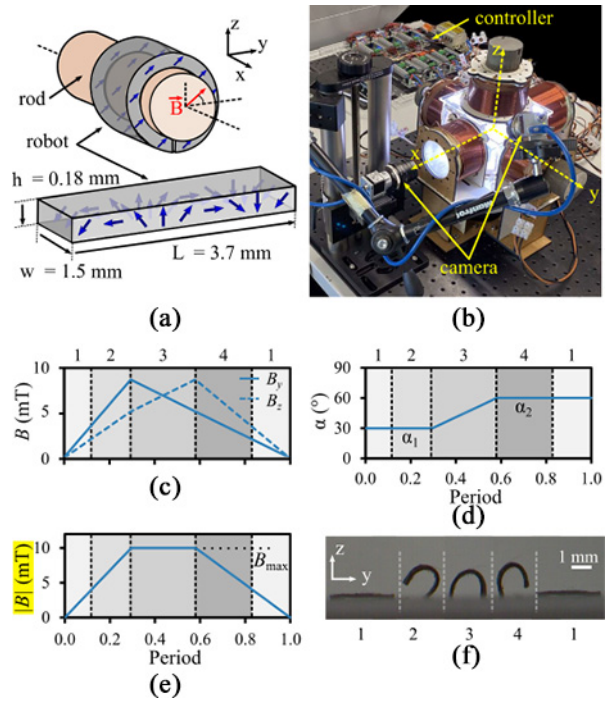

Fig. 1. (a) Fabrication process of the magneto-elastomeric soft millirobot: the robot, which is composed of non-magnetized ferromagnetic microparticles homogeneously distributed inside a silicone elastomer sheet, is rolled around a cylindrical rod and magnetized with $|B|$= 1.8 T field (red arrow) with a 45 angle with respect to the y-axis (see Hu 2018 for details on the fabrication method). The unfolded robot maintains a periodic magnetization profile along its body (blue arrows). (b) Photo of the magnetic actuation and imaging experimental setup with six electromagnetic coils and two high-speed cameras that allow real-time evaluation of the robot's walking gait performance. Walking gait control parameters during a single period of a motion (a sample case of $1/f = 90$ ms and $f = 11$ Hz is normalized to 0–1 on the abscissa). (c) The magnetic field B is controlled on the y–z plane and shown with shown y and z components. (d) The magnetic field orientation changes from α_1 to α_2 and (e) magnitude reaches B_{max}. (f) In a single actuation period of the walking gait, the magnetic soft millirobot follows four consecutive gait states shown with the photos: (1) relaxed, (2) front-stance, (3) double-stance, (4) back-stance, and (1) relaxed again (shown with the numbers above and below the figures).

1. demonstration of a data-driven optimization tool (i.e., BO) that can efficiently learn the gait controllers of a small-scale untethered robot whose performance is prone to fabrication-, material-, and physical interaction-based variabilities;
2. successful testing of the walking gait on three different task spaces that emulate, e.g., dynamic

environments inside the human body, and the adaptation of the robot controller parameters to these environments in a small number of experiments;
3. implementation of an automated experimental platform that runs and evaluates the physical learning experiments repeatably and reliably without human intervention and simulated environments;
4. comparison and evaluation of four different TL methods within the context of GP hyperparameters on the learning efficiency of BO on the small-scale soft robots;
5. generation of five benchmark datasets consisting of the exhaustively parsed controller parameter space involving 3,750 different physical experiments for three different robots and three different walking surfaces that would allow further comparison between different optimization methods, which is available at https://github.com/sozgundemir/softrobotwalkingdataset

The organization of this paper is as follows. We describe the design of the robotic system, the walking gait, and the properties of the task environments in our experiments in Section 2. Section 3 describes the learning approach with the details on the BO, GP, and TL methods. In Section 4, we present the experiments on generating the benchmark datasets, comparing the TL methods, and learning the walking gait in different task environments. We discuss the experimental results and conclude our work in Section 5.

2. Experimental robot system

2.1. Robot design and fabrication

We follow the methods and materials reported in Hu et al. (2018) and our previous work (Culha et al., 2020) and fabricate three magnetic soft millirobots with a 1:1 body mass ratio of Ecoflex 00-10 (Smooth-On Inc.) and neodymium–iron–boron (NdFeB) ferromagnetic microparticles with around 5μ m diameter (MQP-15-7, Magnequench). We place this pre-polymer mixture on a methyl methacrylate plate and cut the robots out of the cast using a high-resolution laser cutter (LPKF Protolaser U4) after the polymer is cured. Our robots have the final dimensions of length $L = 3.7$ mm, width $w = 1.5$ mm, and height $h = 185\mu$ m as shown in Figure 1(a). We separately fold the robots around a cylindrical rod with a circumference equal to L and magnetize them within a magnetic field with a magnitude of 1.8 T and orientation of 45° measured counterclockwise from the y-axis. Once the robots are unfolded from the rod, the magnetic particles maintain their magnetization orientation forming a circular profile along the longitudinal axis of the robot body. We use these robots (i.e., robots 1, 2, and 3), which have the same nominal material properties and dimensions, in the rest of our experiments.

2.2. Walking gait definition

The walking gait of our robot is composed of four consecutive quasi-static states that are inspired by the planar quadrupedal bounding (Alexander, 1984) and a caterpillar inching motion (Trimmer and Lin, 2014). These states are depicted as (1) relaxed, (2) front-stance, (3) double-stance, and (4) back-stance as shown in Figures 1(c)–(f). We control four parameters to generate the walking gait: the maximum magnetic field magnitude ($B_{\max}$), the frequency of the actuation cycle (f), and two magnetic field orientation angles (α_1 and α_2) measured counterclockwise from the y-axis. The plots in Figures 1(c)–(e) show the change of the control parameters during a single period of the motion for $B_{\max} = 10$ mT, $f = 11$ Hz, $\alpha_1 = 30°$, and $\alpha_2 = 60°$, which are adopted from the hand-tuned parameters reported in Hu et al. (2018). At the beginning of a single gait period, the robot starts at a relaxed state for $0 \leqslant |B| \leqslant 4$ mT. The robot tilts forward when $\alpha = \alpha_1$ and $|B|$ increases from 4 mT to $B_{\max} = 10$ mT. Although $|B|$ remains constant at $B_{\max}$, the orientation of the magnetic field changes from α_1 to α_2 causing the robot to initially switch to the double-stance state and then to the back-stance state when $\alpha = \alpha_2$. Then, $|B|$ decreases while keeping the orientation of the magnetic field constant, and the robot gradually switches back to the relaxed state. For $|B| < 4$ mT, the robot assumes the relaxed state, and a single period of walking actuation ends when $|B| = 0$ mT. We reset B at the end of every gait cycle to avoid jerky motion when α changes from α_1 to α_2. In our experiments, the relaxed state is never skipped but its duration changes according to f. The consecutive images from a single walking gait period are shown in Figure 1(f).

2.3. Actuation and feedback setup

We place our magnetized soft robot along the y-axis of the magnetic coil setup consisting of three orthogonal pairs of custom-made electromagnets (see Figure 1(b)) that can generate a 3D uniform magnetic field within a $4 \times 4 \times 4$ cm ^{3} workspace with a maximum value of 15 mT. We modulate the magnetic field on the y–z plane that coincides with the center of the test environment by controlling the electric currents running through the electromagnetic coils with six motor driver units (SyRen25) and an Arduino microcontroller, which runs at 1.2 kHz operation frequency to compute the analog readings received from the current sensors, control the motor drivers and perform all necessary communications with the master PC for each single control cycle of 0.83 ms. We regularly calibrate the magnetic actuation matrix inside the workspace, i.e., the mapping between the applied electric current and the generated magnetic field, to maintain reliable and repeatable experiments.

We track the robot's gait using two high-speed cameras (Basler aCa2040-90uc, shown in Figure 1(b)). The first camera running at 120 frames per second (fps) is placed

orthogonal to the axis of robot motion (i.e., y–z axis of the controller). A tracking algorithm, whose pseudo-code is given in Appendix B, uses this camera to detect and evaluate the robot's motion to identify if the robot is moving according to the walking gait definition given in Section 2.2. The second camera running at 60 fps has an isometric view of the test scene and is used to measure the distance traveled by the robot following the perspective correction of the captured image. In every experiment, we calculate the stride length of the robot by tracking the average distance covered by its center of mass in five consecutive steps. At the end of every experiment, the robot is moved back to its original starting position automatically with the tracking and the actuation commands. See Extension 1 for the gait detection, position tracking, and repositioning for robot 3 walking on paper.

The learning process and image processing run on a master PC, and all the communication tasks between different elements of the robotic system (e.g., image capture and electric current control) are executed on Robot Operating System (ROS) architecture, which allows our system to be scalable for further extensions. The automated experimental platform implemented in our work allows the physical experiments to be executed with minimum human intervention; therefore reducing the human-based disturbances on the robot and the test surfaces. Without these interactions that can cause significant alterations on the soft millirobots, the physical learning experiments can be maintained repeatably and reliably.

2.4. Task environments

In this study, we use a wide range of different task environments to test the efficacy of our adaptive learning strategy in comparison with the limited surface experiments in Culha et al. (2020). Our goal is to emulate the in-air and liquid-immersed surface walking environments that a magnetic soft millirobot might experience during future medical operations inside the human body. To capture some of the characteristic properties of the target tissues and body fluids, we fabricate different task spaces and vary their : (1) surface adhesion, (2) surface roughness, and (3) the liquid medium viscosity properties. For each of these properties, we experimentally identify the range of values that allow successful walking gaits and systematically test the robots in these specific ranges.

We fabricate a set of flat substrates with different surface adhesion strengths by using different materials. This set of substrates consists of paper, polystyrene (PS), and modified polydimethylsiloxane (PDMS). PDMS substrates are prepared by mixing Sylgard184 (Dow Corning) with its curing agent to a 10:1 ratio, degassing, and curing at 90 C for 1 hour. PDMS is modified by adding ethoxylated polyethylenimine (80% solution, Sigma Aldrich) prior to mixing and curing to increase its adhesive properties (Jeong et al., 2016). A volume of 0 (PDMS-0), 1 (PDMS-1), and 2μ; (PDMS-2) of polyethylenimine solution was added per 1 g of silicone elastomer base. The adhesion between the walking surfaces and the robot is measured in a TA Instruments Discovery HR-2 rheometer with a custom-built adhesion setup (at loading and unloading speeds of 50μ m s $^{-1}$). The resulting set of flat substrates resulted in walking surfaces with adhesion strengths between 1 to 10 kPa as shown in Figure 2(a).

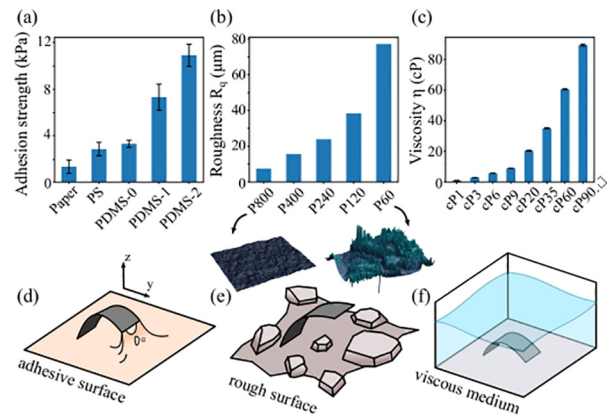

Fig. 2. The type and range of task space properties investigated for the robot's walking gait. (a) Adhesion strength of different surfaces ranges between 1 and 10 kPa. (b) Roughness values of different surfaces that are named after the grit scale of their template sandpapers. The inset figures from the profilometer scans represent the two extremes of the roughness range (i.e., $R_{q,\,P800} = 7.5\mu$ m to $R_{q,\,P60} = 77.2\mu$ m). (c) Viscosity values of the test fluids range between 1 and 90 cP where the robot is submerged while performing the walking gait. Fluids are named after their cP values. The sketches (d)–(f) aim to visualize the task space conditions during walking.

Walking surfaces with changing roughness were prepared by replicating the surface texture of different grits of sandpaper. First, we fabricate the negative molds of the original surfaces by pressing a glass plate with a layer of uncured vinylsiloxane polymer (Flexitime medium flow, Heraeus Kulzer GmbH) onto the original surfaces. After curing the molds for 5 minutes at room temperature, we removed them and produced positive replicas of the original surfaces using clear casting epoxy (EpoxAcast^{TM} 690, Smooth-On Inc., 10:3 ratio by weight) onto the mold, with another glass plate pressed on top. At the end of 24 hours of curing time, we removed the positive replicas from the molds. According to surface profile examination of the replicated surfaces (Keyence VKX260K), surface roughness values R_q (root-mean-square height) changed between 7.5 and 77.2μ m as shown in Figure 2(b).

Last, we submerged the robots in different Newtonian fluids while walking on a flat surface to investigate the effect of bulk liquid medium viscosity on the walking performance. The Newtonian fluids were prepared by mixing different ratios of water and glycerol, and their viscosity is measured in a TA Instruments Discovery HR-2 rheometer. We analyze task environments with medium viscosity ranging from 1 to 90 cP as shown in Figure 2(c).

3. Learning approach

We adapt the learning approach from our previous work (Culha et al., 2020) that aims to optimize the walking gait controller parameters to maximize the stride length S of the robot. Here we define the reward function as

$$S : \Theta \rightarrow R \tag{1}$$

which maps the parameter set $\theta = [B_{\max}, f, \alpha_1, \alpha_2]$ to scalar reward values (i.e., the stride length performance of a robot). According to the definition of the reward function, we formulate the parameter learning as the (global) optimization problem

$$\theta^* = \underset{\theta \in \Theta}{\operatorname{argmax}}\, S(\theta) \tag{2}$$

where Θ denotes the complete search space, θ is the parameter set, and $S(\theta)$ is the average stride length performance of the robot for a given θ.

We define the range of the controller parameters based on the findings in Hu et al. (2018) and the physical limitations of our magnetic actuation setup. Accordingly, $B_{\max}$ is defined between 5 and 12 mT, and the walking frequency, f, ranges from $f_{\min} = 0.5$ Hz to $f_{\max} = 20$ Hz. We limit α_1 and α_2 to $[0, 70]^\circ$ and $[30, 90]^\circ$, respectively, and select values that satisfy $\alpha_2 > \alpha_1$ to generate the walking gait in Figures 1(c)–(f). We use a step size of 1 mT for $|B|$, 5° for each α, and a variable step size of 0.5 Hz for $f < 2$ and 2 Hz for $f \geqslant 2$ Hz, which yield a total number of 15,600 possible parameter values in Θ.

3.1. Gaussian processes

The magnetic soft millirobots in our paper did not have accurate models for kinematics or dynamics (i.e., we demonstrated the model inaccuracy of the original work of Hu et al. (2018) in our previous study in Culha et al. (2020)), therefore, it is necessary to approximate the reward function based on the data collected from physical experiments. However, the physical data has inherent uncertainty owing to the noise in the measurements and the variations during the experiments. To include these uncertainties in the model, overcome the sparsity in the data, and make probabilistic predictions at unobserved locations, we represent the reward function $S(\theta)$ using GPs following the study in von Rohr et al. (2018):

$$S(\theta) \sim GP(\mu(\theta), k(\theta, \theta')) \tag{3}$$

where $\mu(\theta)$ is the prior mean and $k(\theta, \theta')$ is the kernel function defining the covariance between $S(\theta)$ and $S(\theta')$ for $\theta, \theta' \in \Theta$. However, as $S(\theta)$ can only be measured with noise, we define the observed stride length $\tilde{S}$ as

$$\tilde{S}(\theta_i) = S(\theta_i) + n_i \tag{4}$$

where n_i is zero-mean Gaussian noise with variance σ_{n}^2 for each measurement i.

During one run of BO, the GP model is sequentially updated with $\tilde{S}(\theta)$ observed from experiments. We define one *learning run* as a run of BO until the desired stopping criterion is satisfied (e.g., a fixed number of experiments is reached).

From the experimental data $D = \{\theta_i, \tilde{S}(\theta_i)\}_{i=1}^{N}$, where N denotes the number of experiments in D, the stride length of the robot for an unobserved θ can be predicted using the posterior mean and variance as

$$\mu_{\mathrm{post}}(\theta) = \mu(\theta) + k^T(\theta) K^{-1} y \tag{5}$$

$$\sigma_{\mathrm{post}}^2(\theta) = k(\theta, \theta) - k^T(\theta) K^{-1} k(\theta) \tag{6}$$

$$S_{\mathrm{post}}(\theta) | D \sim \mathcal{N}(\mu_{\mathrm{post}}(\theta), \sigma_{\mathrm{post}}^2(\theta)) \tag{7}$$

where $k(\theta), y \in R^N$ with $[k(\theta)]_i = k(\theta, \theta_i)$, $y_i = \tilde{S}(\theta_i) - \mu(\theta_i)$, and $K \in R^{N \times N}$ with $K_{i,j} = k(\theta_i, \theta_j) + \delta_{i,j}\sigma_{\mathrm{n}}^2$, where $\delta_{i,j}$ is the Kronecker delta.

We select the squared exponential as the kernel function in the GPs, which is defined in (Duvenaud et al., 2011) for multi-dimensional cases as

$$k_{\mathrm{SE}}(\theta, \theta') = \sigma_{\mathrm{f}}^2 \exp\left(-\sum_{u=1}^{d_c} \frac{(\theta^u - \theta'^u)^2}{2(l_{\mathrm{c}}^u)^2} \right) \tag{8}$$

where $l_{\mathrm{c}} \in R^{d_c}$ is the length scales that defines the rate of variation in the modeled function for each dimension of the parameter space, i.e., l_{c}^u for $u \in \{B_{\max}, f, \alpha_1, \alpha_2\}$. Roughly speaking, long length scales are used to model slowly-varying functions and short length scales are used to model quickly-varying functions. The signal variance σ_{f}^2 describes the width of distribution, e.g., high σ_{f}^2 means higher uncertainty in the predictions of the unobserved θ. We implement the GP model in our experiments using the libraries provided by GPy (GPy, 2012).

3.2. Bayesian optimization

We use BO to select the parameter set θ_{next} to be tested in the next step of the learning run using the acquisition function $\alpha_{\mathrm{acq}}(\theta)$ as

$$\theta_{\mathrm{next}} = \underset{\theta \in \Theta}{\operatorname{argmax}}\, \alpha_{\mathrm{acq}}(\theta). \tag{9}$$

In this study, we choose the expected improvement (EI) as the acquisition function $\alpha_{\mathrm{acq}}(\theta)$ due to its better performance compared with its alternatives as demonstrated in von Rohr et al. (2018). EI seeks the parameter set for the next step where the EI in reward function is the highest compared with the previously collected data and is defined in Jones et al. (1998) as

$$\alpha_{\mathrm{acq}}(\theta) = \mathbb{E}[\max(0, (S(\theta) - \tilde{S}(\theta^*)))] \tag{10}$$

where $\tilde{S}(\theta^*)$ is the highest reward function value collected so far. Analytical solution of Equation (10) is given in Brochu et al. (2010) as

$$\alpha_{\text{acq}}(\theta) = (\mu(\theta) - \tilde{S}(\theta^*) - \xi)\Phi(Z) + \sigma(\theta)\phi(Z) \quad (11)$$

where Φ and ϕ are the Gaussian cumulative density and probability density functions, respectively. The term Z is described as $Z = Z(\theta) = (\mu(\theta) - \tilde{S}(\theta^*) - \xi)/\sigma(\theta)$, with $\mu(\theta)$ and $\sigma(\theta)$ computed from Equations (5) and (6). The two terms in Equation (11) define the exploitation and the exploration weights of the BO, respectively. The balance between these two terms is controlled by the hyperparameter ξ. As ξ gets higher, BO focuses more on exploration and seeks the next parameter set in regions with high prediction uncertainty. In contrast, BO focuses more on exploitation and selects the next parameter set within a close range to already explored regions. As the goal of our study is to adapt to task spaces by increasing the likelihood of finding more controller parameter sets that yield successful walking gaits under uncertainty, we choose $\xi = 0.1$ to increase the exploration tendency of the BO in our experiments.

3.3. Transfer learning

In this study, we compare four different methods of TL on our walking gait experiments: (1) transfer of all GP hyperparameters, σ_{n}^2, σ_{f}^2, and length scales l_{c} for each dimension of the parameter space $\mathbb{R}^{d_c}$, i.e., l_{c}^u for $u \in \{B_{\max}, f, \alpha_1, \alpha_2\}$; (2) transfer of only the length scales l_{c}; (3) transfer of prior mean information $\mu(\theta)$; and (4) the hybrid combination of length scales l_{c} and $\mu(\theta)$.

3.3.1. Transfer of GP hyperparameters. The choice of the types and values of GP hyperparameters influence the regression of the GP (Chen and Wang, 2018) and their transfer from prior models can change the dynamics of the learning process (Patacchiola et al., 2020; Wang et al., 2020). The hyperparameters we choose to investigate as a part of the GPs in this study can be listed as the noise in the collected data σ_{n}^2, the signal variance σ_{f}^2, and the length scales l_{c}. We start the BO learning by initializing the σ_{n}^2 to the maximum variance found in the repeated experimental results in our previous work in Culha et al. (2020), and setting the signal variance σ_{f}^2 to the square of half of the body length of the robot so that the highest possible reward value (i.e., $L = 3.7$ mm) remained inside the 95% confidence interval of the prior. We also set the length scale values l_{c}^u to a quarter of the total range of each corresponding parameter. After starting the BO runs with these initial values, we use the *log marginal likelihood estimation* (Rasmussen and Williams, 2006)

$$\log p(\tilde{S}|\theta, l_{\text{c}}^u, \sigma_{\text{n}}^2, \sigma_{\text{f}}^2) = -\frac{1}{2}(\tilde{S} - \mu(\theta))^{\text{T}} K^{-1} (\tilde{S} - \mu(\theta)) - \frac{1}{2}\log|K| - \frac{N}{2}\log 2\pi \quad (12)$$

to simultaneously optimize the GP hyperparameters based on the collected data during the learning runs. We use these estimates of the selected hyperparameters as one of the TL methods in the following experiments in Section 4.3.

3.3.2. Transfer of prior mean information. In addition to the kernel, the prior mean $\mu(\theta)$ must be chosen at the beginning of a BO run as well. Often, constant zero mean (i.e., $\mu = 0$) is the default choice as an uninformed prior mean function for maximization problems (Chen and Wang, 2018). For the millirobot learning problem herein, we investigate the transfer of information from previous learning runs by setting the prior mean to the posterior mean of a previously trained GP model, such as from a different robot. In this way, we can approximately transfer the topology of the target function between different test scenarios, which is reasonable as long as the differences between the robots and the environments do not significantly alter the function shape.

3.3.3. Hybrid transfer. Previous methods can be combined and both the optimized estimation of the GP hyperparameters and the prior mean information can be transferred between the BO experiments. In this study, we also investigate the combination of the estimated length scales l_{c} and the prior mean information $\mu(\theta)$ and their transfer between the test cases in Section 4.3.

4. Experimental results

Our study aims to use BO and GPs to demonstrate adaptation to different task spaces while experimentally optimizing the stride length of the soft millirobots whose walking performances are prone to fabrication-, material-, and interaction-based reproducibility issues that cannot be successfully predicted with kinematic models. In that sense, we focus more on exploring a variety of walking patterns under changing task space conditions rather than continuously optimizing a specific walking gait performance. Accordingly, we design the experiments to highlight the influence of BO and GPs and TL methods on increasing the average performance of finding successful walking gaits, i.e., gaits strictly following the consecutive states described in Section 2.2 that also yield sub-optimal stride length performances, during the limited number of learning runs, instead of only finding the optimum controller parameters.

We begin with using an exhaustive search approach to generate benchmark datasets for the walking gaits on five different test scenarios using our millirobots in Section 4.1. Here, we limit the controller parameter space to two dimensions and only explore the α_1 and α_2 parameters while experimenting with three robots on a flat paper surface and with one robot (i.e., robot 3) on two additional different walking surfaces. The results of the exhaustive

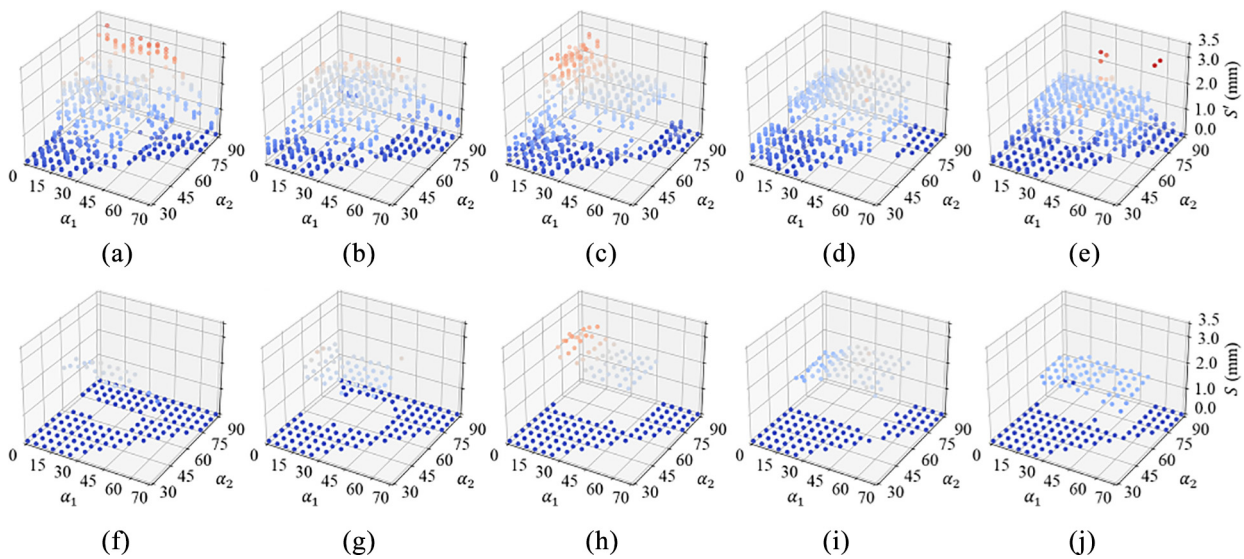

Fig. 3. Experimental displacement measurements S' in five test case scenarios for (a) robot 1 on paper, (b) robot 2 on paper, (c) robot 3 on paper, and robot 3 on (d) PDMS-0 and (e) P800 surfaces. Each figure represents 5 repetitions for 150 different controller value pairs, yielding 750 physical experiments. The displacement measurements are filtered out with the tracking algorithm to identify the desired walking gaits and the corresponding average stride length S values for the benchmark datasets (f) robot 1 on paper, (g) robot 2 on paper, (h) robot 3 on paper, and robot 3 on (i) PDMS-0 and (j) P800 surfaces.

search show the overall structure of the walking gait function based on the range of the two controller parameter values. This statistical information serves as benchmark data to compare the learning efficacy of the BO and different TL methods in Sections 4.2 and 4.3. We choose the most effective TL method guided by the experiments on the benchmark datasets and apply it to a new set of task adaptation experiments in Section 4.4. Here we use a single robot on three different task spaces with a wide range of changing surface adhesion and roughness, and medium viscosity. In these experiments, we expand the controller space back to four dimensions (i.e., $B_{\max}$, f, α_1, and α_2), and optimize the walking gait with the BO and the chosen TL method within a limited number of learning runs.

4.1. Generation of the walking gait benchmark datasets

In our previous work in (Culha et al., 2020), we observed that the soft millirobots we adopted from (Hu et al., 2018) experienced additional material degradation over long repeated experiments that altered their gait performances. While we investigate the influence of BO and TL methods on the improvement of learning efficiency in Section 4.2 and Section 4.3, we want to minimize this material degradation effect on the walking gait. That is why here we use an exhaustive search approach and generate five different benchmark datasets that cover the walking gait function space necessary for the BO and TL methods investigations. To this end, we test three different robots (i.e., robot 1, 2, and 3) on a flat paper surface and a single robot (robot 3) on two additional surfaces: PDMS-0 and P800-grit sandpaper replica.

To explore the walking gait function space on these five test cases, we constrain the controller space into two dimensions by using a constant $B_{\max} = 10$ mT and $f = 1$ Hz, and changing the α_1 and α_2. We choose values from the set $[0, 70]^\circ$ for α_1 and $[30, 90]^\circ$ for α_2 with a step size of 5° that meet the condition $\alpha_2 > \alpha_1$, which consequently generate 150 different controller parameter pair values. For each of these pairs, we repeat the experiments five times; hence, generating 750 physical experiments for each test case and report the results in Figures 3(a)–(e). The constrained dimensions reduce the necessary experiments from 390,000 (i.e., 5 repetitions for each test case using 15,600 parameter value sets) to 3,750 for the exhaustive search, and significantly avoid the possible material-degradation over prolonged experiments.

In these experiments the robots do not necessarily follow the gait definition in Section 2.2, therefore we describe these resulting values as ''displacement measurements,'' S'. To filter out the non-walking gaits of the robot from this dataset, which can be seen at the beginning of Extensions 2, 3, and 4 for different task spaces, we use our gait tracking feedback system defined in Section 2.3 and evaluate every test result to penalize the α controller pairs that do not generate the desired walking gait. Accordingly, we obtain the average stride length performances, S, of the successful gaits in Figures 3(f)–(j). These results show that the stride length performances are limited up to $S \approx 2.4$ mm for the successful walking gaits, and the penalized motions with higher displacements ($S' > S$) do not comply

Table 1. Best performing α controller parameter sets found by the exhaustive search and the corresponding stride length results $\overline{S_{\text{exh}}}$

Test case	Controller parameters*		Stride length $\overline{S_{\text{exh}}}$ (mm) (average ± SD)
	$\alpha_1(^\circ)$	$\alpha_2(^\circ)$	
Robot 1 on Paper	15.0	70.0	1.75 ± 0.04
Robot 2 on Paper	0.0	70.0	1.91 ± 0.12
Robot 3 on Paper	5.0	70.0	2.43 ± 0.12
Robot 3 on PDMS-0	30.0	85.0	1.83 ± 0.09
Robot 3 on P800	35.0	70.0	1.77 ± 1.37

*$B_{\max} = 10$ mT and $f = 1$ Hz in all cases.

with the walking gait definition. The statistical information collected from the physical experiments in these five test cases (shown in Figures 3(f)–(j)) constitutes our benchmark datasets that we use in the following BO and TL investigations. The optimum controller parameter sets found by the exhaustive search in the 2D function space are reported in Table 1. We use the mean and standard deviation values of these exhaustive search results to sample the stride length of the robots for the given α controller parameter sets while comparing the performances of different TL methods. The benchmark data is available online and can be accessed at https://github.com/sozgundemir/softrobotwalkingdataset.

4.2. Learning the walking gait with the "standard" BO

We initially test the walking gait learning with a BO approach on the benchmark datasets, where the prior mean information is set to zero (i.e., $\mu(\theta) = 0$) and the four hyperparameters are set to initial values described in Section 3.3.1 without inheriting any other prior information. We utilize this approach, which we refer to as "standard BO" in the rest of this study, for the two controller parameters α_1 and α_2 and apply it separately to the benchmark datasets for the five different test cases (i.e., robots 1–3 on a flat paper surface, and robot 3 on the PDMS-0 and P800 surfaces). To be able to directly use the mean and standard deviation information from the exhaustive search results, we configure our BO to explore the same discrete controller parameter set space used in Section 4.1. We perform 100 independent learning runs with each involving 100 iteration steps for the five test cases. One iteration step of a learning run involves three steps:

1. BO selects a new parameter set θ that maximizes the acquisition function based on the GP model;
2. for the selected controller parameter pair, the corresponding stride length performance is sampled from the normal distribution defined by the mean and standard deviation values found in the relevant benchmark dataset;
3. the learning system updates the GP model using this sampled data and prepares for the next iteration step of the learning run.

We report the median of the learning results with the upper and lower quartiles in Figure 4. These values represent the normalized gait performance $\hat{S} = \tilde{S}/\overline{S_{\text{exh}}}$, where $\overline{S_{\text{exh}}}$ is the mean of the stride length performances of the robots for the best α controller parameters reported in Table 1. For the first four of the five test cases, the standard BO approach finds the optimum gait controller parameters in less than an average of 25 iterations out of 100 independent learning runs. For the robot 3 walking on the P800 surface, the BO finds approximately 74% ($\hat{S} \approx 0.74$) of the optimum stride length performance as shown in Figure 4(e). The results show that the standard BO starts the learning without any prior information and occasionally finds controller parameter sets that yield $\hat{S} = 0$ in the first 15 iteration steps. The non-monotonic optimization results in these initial steps are expected due to the statistical exploration nature of our BO approach. After 25 iteration steps, the BO maintains the exploration in the close vicinity of the optimum controller parameters it finds so far for all the test cases (i.e., $\hat{S} \approx 1$ for robots 1–3 on paper and robot 3 on PDMS-0, and $\hat{S} \approx 0.74$ for robot 3 on P800). The variation of the generated walking gait performances is bounded by the standard deviations reported in Table 1. The exact normalized performance results for the standard BO approach can also be seen in Table 2.

4.3. Comparison of transfer learning methods on the benchmark datasets

In this study, we extend our previous investigation on the role of TL in learning efficiency (Culha et al., 2020) and compare four different methods while optimizing the gait controllers of our soft millirobots. Similar to Section 4.2, we apply our BO learning to the benchmark datasets generated in Section 4.1, where the controller parameter space is limited to two dimensions with α_1 and α_2. However, unlike the "standard" BO, here we initialize the learning runs of the robots in all test cases with different types of

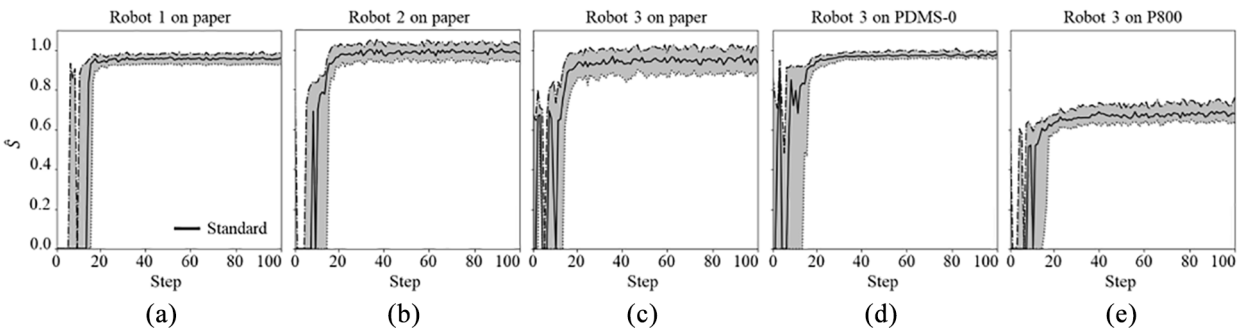

Fig. 4. Performance of the standard BO for (a) robot 1 on paper, (b) robot 2 on paper, (c) robot 3 on paper, (d) robot 3 on PDMS-0, and (e) robot 3 on P800. The stride length performances are normalized with respect to $\overline{S_{\text{exh}}}$ given in Table 1 for easier comparison between tests. Each figure shows the statistical results as median, and upper and lower quartiles from 100 independent BO runs with each consisting of 100 iterations.

Table 2. Comparison of the transfer learning methods.

Robot Number	Test Surface	Type	Performance *(%)	Performance **(%)	Convergence Step
1	Paper	Standard	95.83	↔	20
		HP-4	95.35	−0.51	9
		HP-l_c	96.17	+0.35	28
		Mean	97.20	+1.42	2
		Hybrid	97.34	+1.57	3
2	Paper	Standard	99.63	↔	21
		HP-4	86.32	−13.36	100
		HP-l_c	99.62	−0.01	29
		Mean	98.44	−1.19	3
		Hybrid	99.40	−0.23	6
3	Paper	Standard	95.60	↔	20
		HP-4	81.25	−15.02	100
		HP-l_c	98.03	+2.54	27
		Mean	98.81	+3.36	0
		Hybrid	98.23	+2.75	0
	PDMS-0	Standard	98.38	↔	25
		HP-4	86.90	−11.68	100
		HP-l_c	97.74	−0.66	36
		Mean	96.54	−1.87	14
		Hybrid	96.68	−1.73	8
	P800	Standard	73.84	↔	30
		HP-4	77.61	+5.11	14
		HP-l_c	80.00	+8.33	30
		Mean	91.12	+23.40	8
		Hybrid	84.47	+14.40	12

*Relative performance with respect to the optimum exhaustive search results.
**Relative performance with respect to the standard BO learning approach.

prior information learned from the robot 3 walking on a flat paper surface. In the remaining parts of this section, we refer to this prior information source as the *source robot*. The four different TL methods we compare are: (1) transfer of all GP hyperparameters (*HP-4 transfer*), (2) transfer of only the length scales (*HP-l_c transfer*), (3) transfer of prior mean information (*mean transfer*) and (4) the hybrid combination of length scales with mean information (*hybrid transfer*). All the experiments shown in this section are the result of 100 independent learning runs with each having 100 iterations for all the test cases. The values for the hyperparameters used in each of the TL methods can be found in Appendix C.

4.3.1. HP-4 transfer. We initially transfer all of the four GP hyperparameters (i.e., noise in the collected data σ_n^2, signal variance σ_f^2, and length scales l_c^u for $u \in \{\alpha_1, \alpha_2\}$) that are optimized with the log marginal likelihood estimation in (12) from the learning runs of the source robot to all other five test cases. Here, we initialize the BOs with this prior information and start the learning experiments. The normalized stride length performances, $\hat{S}$, of the BO learning with this TL method (depicted as "HP-4") are shown in comparison with the "standard" BO approach in the first column of Figure 5. These results show that the transfer of all of the four hyperparameters improves the learning performance in terms of decreasing the number of iteration steps to find the optimum controller parameters within these experiments. For robot 1 on paper and robot 3 on P800, the BO manages to find the performances achieved by the "standard" approach in less than half of the iteration steps, which are shown in Figure 5(a) and (q), respectively. For robots 2 and 3 on paper, and robot 3 on PDMS-0 (as seen in Figure 5 (e), (i), and (m)), the BO with this TL method does not reach the performances previously achieved by the "standard" BO. After finding the optimum performances, the BO maintains the exploration in their close vicinity for all the test cases.

4.3.2. HP-l_c transfer. Second, we investigate the transfer of only the two length scale hyperparameters l_c^u for $u \in \{\alpha_1, \alpha_2\}$ optimized by the source robot. We run the BO learning experiments after all the length scale parameters are initialized with the transferred l_c values. The results in the second column of Figure 5 show that this TL method (depicted as "HP-l_c") worsens the learning performance as it increases the number of iteration steps for the BO to find the optimum performance parameters for

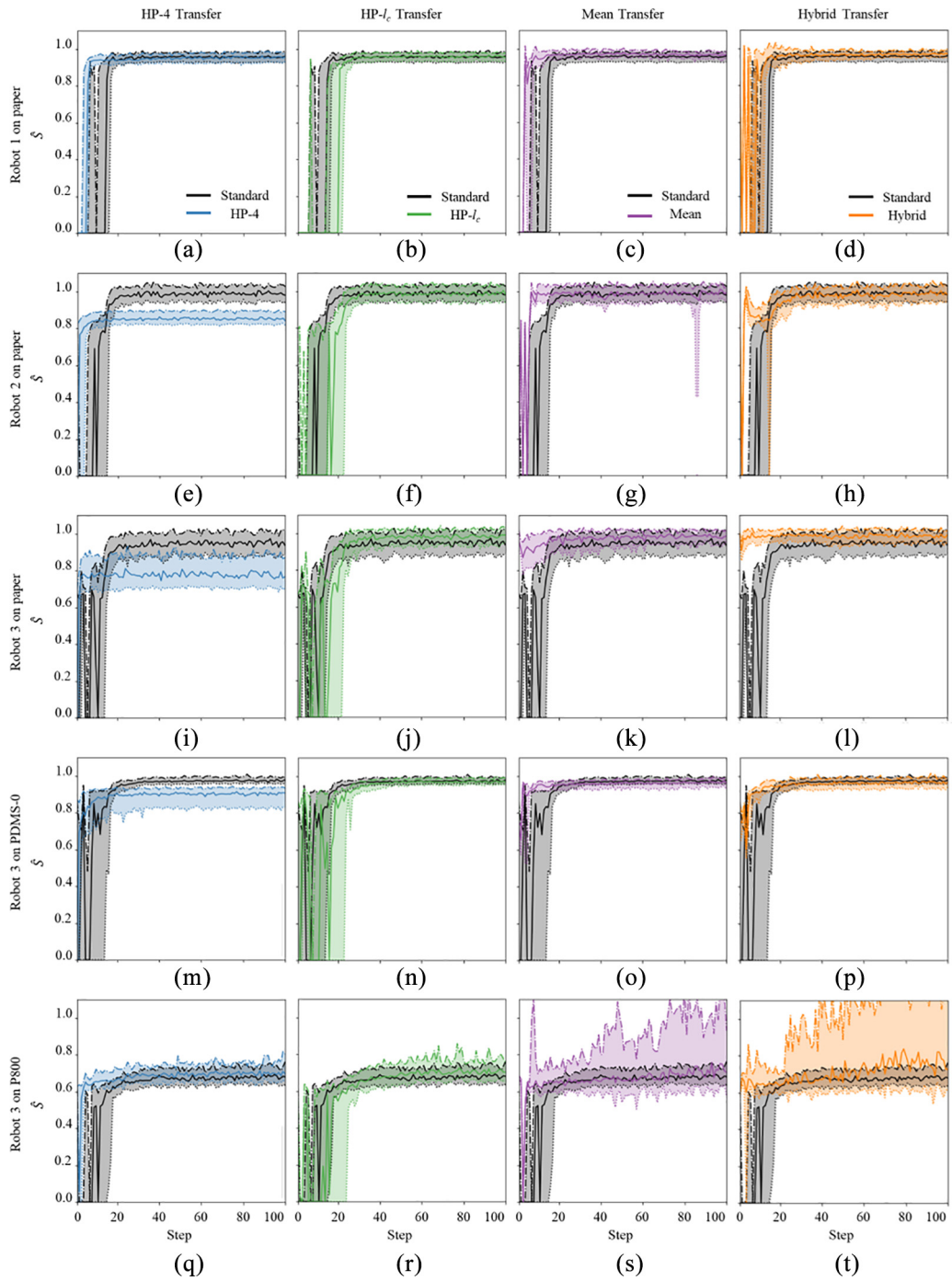

Fig. 5. Comparison of each TL method are shown in separate columns in terms of normalized stride length performance $\hat{S}$ with the standard BO approach on five test cases (first row) robot 1 on paper, (second row) robot 2 on paper, (third row) robot 3 on paper, (fourth row) robot 3 on PDMS-0, and (fifth row) robot 3 on P800. The grayscale plots represent the standard BO replicated from Figure 4. Straight lines represent the median and the shaded regions between dashed lines show the upper and lower quartiles. Each figure shows the statistical results from 100 independent BO runs with each consisting of 100 iterations.

all the cases. However, in comparison with the previous "HP-4" method, the transfer of only the length scale hyperparameters allows the BO to explore the gait performances achieved by the standard BO approach.

4.3.3. Mean transfer. As the third method, we transfer the posterior mean information, $\mu(\theta)$, from the source robot as the prior information for the robots in the test cases. Here, we set all of the other GP hyperparameters to the initial

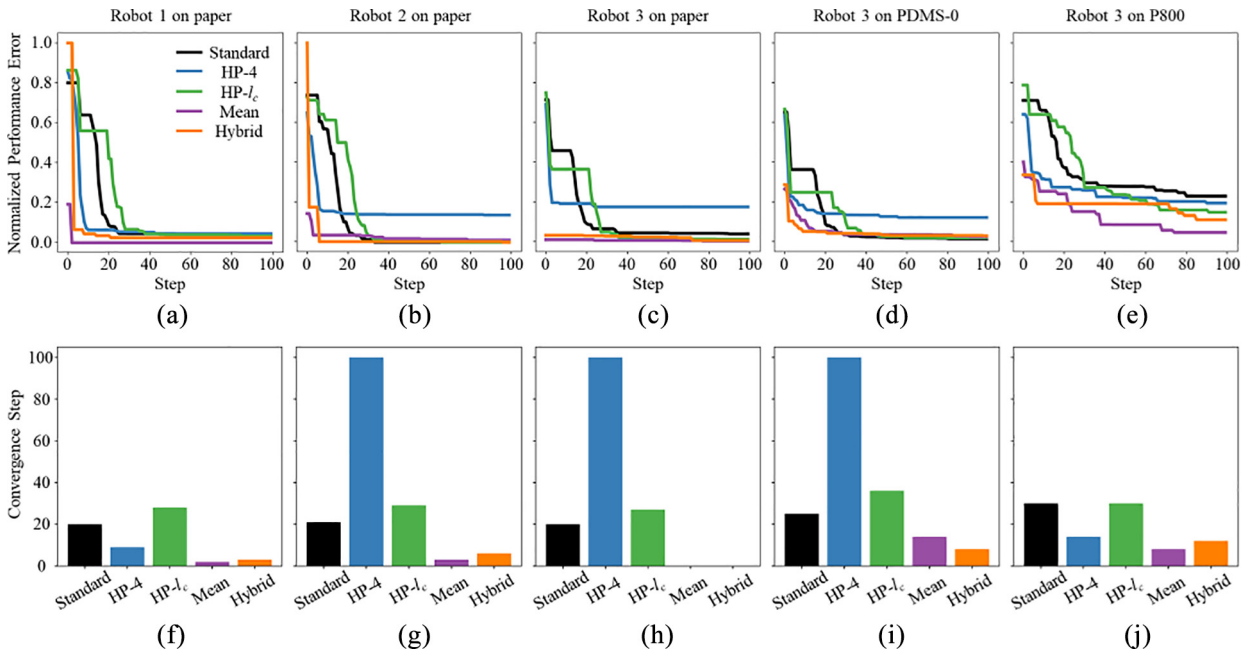

Fig. 6. Performance comparison of standard BO and four TL methods on five test cases. (a)–(e) Normalized performance error ε with respect to the optimum gait performances from the exhaustive search results (Table 1). Each figure shows the "best-so-far" performance results over 100 iterations. (f)–(j) Comparison of resulting convergence step of standard BO and TL methods. The convergence steps of the TL methods are calculated as the iteration step achieving the performance level of the standard BO approach. Maximum iteration step of 100 is reported for the cases that cannot reach the standard BO level.

values as described in Section 3.3.1 and restart the BO learning experiments. The results in the third column of Figure 5 show that when the BO learning starts with the prior mean information (depicted as "Mean"), it achieves the average gait performances found by the standard BO ($\hat{S}\approx 1$) in fewer iteration steps for the first four test cases. Although the median of the achieved $\hat{S}$ remains close to the standard BO results for robot 3 on P800, this method also allows the exploration of the controller parameter sets those yield performances close to the optimum results from the exhaustive search as shown in Figure 5 (s).

4.3.4. Hybrid transfer. Finally, we adopt a hybrid approach and transfer the posterior mean information, $\mu(\theta)$, together with the two length scale hyperparameters l_c^u for $u \in \{\alpha_1, \alpha_2\}$, whose results are shown in the last column of Figure 5 (depicted as "Hybrid"). The hybrid transfer improves the learning run performance by decreasing the number of iteration steps to find the optimum performing parameter sets found by the standard BO. Similar to the results in Figure 5 (s), the hybrid transfer also allows robot 3 on P800 to investigate parameter regions that yield performances close to the optimum results from the exhaustive search.

The comparative performance results of the standard BO and the four TL methods for each of the five test cases are reported in Figure 6 and Table 2. Owing to the statistical and explorative nature of the BO, the stride length performances do not monotonically increase at every consecutive iteration step (as also visible in Figure 5). That is why, to provide a clear comparison between our methods in Figure 6, we identify the iteration steps that show the "best so far" performance during the learning run of each approach. The first row, Figure 6(a)–(e), compares these methods in terms of the normalized error of the achieved stride length performances during the BO learning runs ($\varepsilon = 1 - \hat{S}$). For the first four cases, it can be seen that except for the "HP-4" method, the other three transfer methods and the standard BO manage to consistently explore the optimum stride length performances ($\varepsilon\approx 0$) as shown in Figure 6(a)–(d). The "HP-4" method does not allow the BO to find the optimum controller parameters and the results remain approximately 20% below these optimum values for three of the test cases, which are also shown in Figure 6(b)–(d). The "HP-$\mathbf{l}_c$" method shows similar performances with the standard BO in terms of the initial and the final performance error. Both the "Mean" and "Hybrid" transfer methods allow the BO to generate comparable final performances, while the mean transfer method typically starts with significantly lower error compared with the hybrid approach. For the last test case (i.e., robot 3 on P800, Figure 6(e)) we can see that the standard BO approach only manages to explore gaits that are 74% ($\hat{S}\approx 0.74$) of the optimum gait performance. Every TL method increases the performance yield of the BO learning, with the "Mean" transfer outperforming all other methods by exploring gaits that have $\hat{S}\approx 0.91$, or $\varepsilon\approx 0.09$, on the surface of P800-grit sandpaper

in terms of the final performance. The comparison between the standard BO and the TL methods in terms of achieved $\hat{S}$ performances are given in detail in Table 2.

The second row, Figure 6(f)–(j), represents the learning efficiency performance comparison in terms of the iteration steps needed to explore the best performance by the standard BO and to achieve standard BO level performance by the four TL methods. We describe this exploration performance with ''convergence steps,'' which is calculated by finding the performance value that stays within the 5% band of the averaged remaining steps. Normally, a monotonic convergence is not expected from the statistical and explorative BO learning. However, viewing the learning runs with the ''best-so-far'' evaluation method allows us to represent the required iteration steps to achieve comparable performance, and to capture the relative data efficiency of the TL methods. Accordingly, we can see that the ''HP-4'' method fails to achieve standard BO level performance for the three cases (Figure 6(g)–(i)) as its convergence steps are equal to the number of iteration steps in the experiments. The ''HP-$\mathbf{l}_c$'' method shows comparable results with the standard BO for all the cases (Figures 6(f)–(j)) in terms of the convergence steps. In comparison, both the ''Mean'' and ''Hybrid'' TL methods attain to the standard BO level performance significantly faster (i.e., fewer convergence steps) in all the test cases. The details of the convergence steps is given in the last column of Table 2.

As the ''Mean'' TL method outperforms the standard BO and other three TL methods by finding better performing parameter sets in fewer iterations, we select it to use for task space adaptation experiments in Section 4.4.

4.4. Adaptation to task spaces

The task environment is more susceptible to dynamic changes than the robot morphology, especially for medical operations inside the human body. Therefore, a quick adaptation of the robot controller is important to maintain successful robot task handling. In the following experiments, we investigate the learning efficiency of our BO approach while focusing on the three physical properties that may dynamically change during the walking task of our soft millirobots in future *in vivo* operations, which are (1) surface adhesion, (2) surface roughness, and (3) medium viscosity.

Here, we expand the controller parameter space exploration back to four dimensions by including the magnetic field magnitude $B_{\max}$ and the actuation frequency f. To efficiently learn the controller parameters in this higher dimensional search space, we utilize the prior mean transfer application (i.e., ''Mean'' TL), which is shown to be the best performing TL method for our experimental scenario in Section 4.3. However, as the source robot in previous experiments explored only two dimensions of the controller space, we generate the posterior mean $\mu_{\text{post}}(\theta)$ required for the following experiments by performing a new set of physical experiments with the robot 3 on a flat paper surface. For these experiments, we run the standard BO for 156 iteration steps (i.e., 156 different controller parameter sets), which corresponds to 1% of the complete controller parameter space with given step sizes in Section 3. We use the posterior mean optimized at the end of these learning runs as the prior mean information for all the task space experiments. We only test robot 3 in the following task space adaptation experiments.

Again, we compare the learning efficiency of the standard BO with the prior mean transfer method on all the task spaces defined in Section 2.4. Here, the objective of these learning runs is to adapt to dynamic task spaces and learn the optimized controller parameters in as few experiments as possible especially for future medical operations. To find the number of learning steps sufficient enough for BO to find the desired walking gaits, we used the results in Figures 4, 5, and 6, which involve approximately 250,000 data points. These results show that the BO finds the optimized controller parameters that generate desired walking gaits consistently in less than 20 steps for different robots and walking surfaces. Therefore, we limit the number of steps of a learning run to 20 experiments (i.e., iteration steps), and perform three independent learning runs with the same initial conditions, yielding 60 experiments in total. One step of the learning run involves five steps:

1. BO selects a new parameter set θ that maximizes the acquisition function based on the GP model;
2. the microcontroller initiates the physical experiment and regulates the magnetic field based on the selected θ;
3. the cameras record the robot's motion and measure the average stride length performance $\tilde{S}$ after running the gait tracking system;
4. the learning system updates the GP model using the newly collected data from the experiment;
5. the robot returns to its initial position for the next experiment.

4.4.1. Surface adhesion. We initially test the robot on five surfaces with different adhesion strengths reported in Figure 2(a). Figure 7(a)–(d) show the walking gait performances during the three independent learning runs on the two ends of the adhesion range: paper (1.34 kPa) and PDMS-2 (11.02 kPa). The left column (Figure 7(a) and (c)) show the learning runs with the standard BO approach, and the right column (Figure 7(b) and (d)) show the learning runs with the prior mean transfer method. The difference between these figures shows that the TL method improves the learning runs by finding more of the controller parameters that yield positive walking gait performances. In addition, the BO with the TL manages to explore these parameters in the earlier steps of the learning runs compared to a standard BO approach. We represent

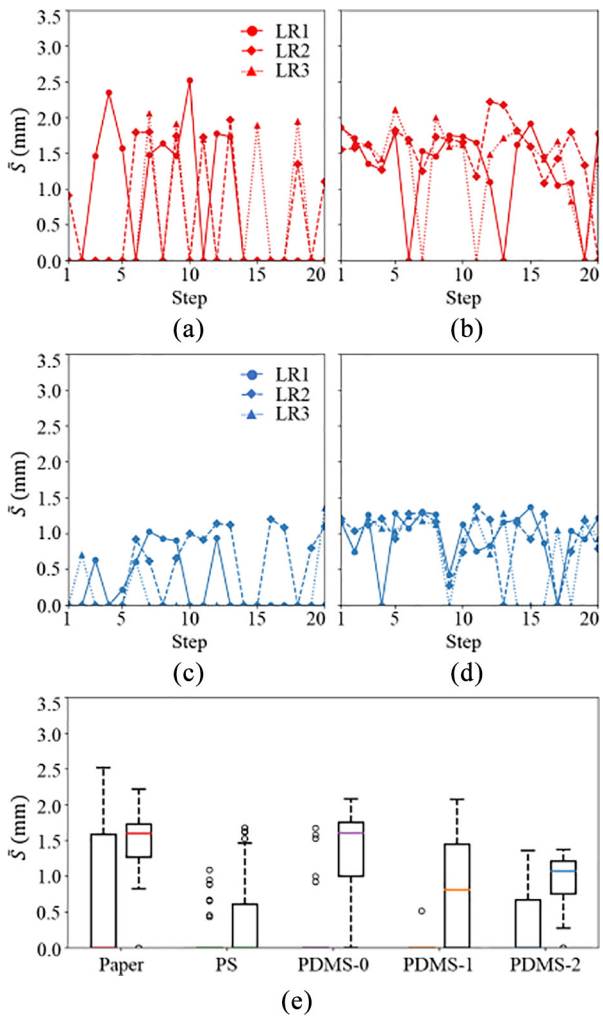

Fig. 7. The learning of the controller parameters for the changing adhesion properties of the test surfaces (upper row paper, lower row PDMS-2) within 20 physical experiments in 3 independent learning runs (depicted as LR 1–3). Learning runs with the standard BO in (a) and (c) are compared with the learning runs with the mean transfer in (b) and (d). Overall performances of the learning runs with the standard BO (left bars) and the mean transfer method (right bars) reported with box plots for all test cases (e).

the general influence of the mean transfer method on all the adhesion surfaces with the standard interquartile range (IQR) method in Figure 7(e). In this figure, the horizontal lines represent the median of the generated $\tilde{S}$ in 60 experiments for each surface. These lines are surrounded by boxes that show the upper and lower quartiles. The error bars show the extremes in terms of the highest and lowest $\tilde{S}$ performances and the circles represent the outlier $\tilde{S}$ performances. The results for the learning with standard BO and BO with the chosen TL method are given side by side for each test surface in Figure 7(e) and the exact values are listed in Table 3. For example, we can see that the learning runs with the standard BO on PS, PDMS-0, and PDMS-1 surfaces typically explore controller parameters that do not yield successful walking gaits. In comparison, the BO with the TL method explores numerous successful controller parameters on these surfaces. The increased median lines in Figure 7(e) provide evidence that the prior mean transfer improves the learning of the BO in terms of increasing the number of successful walking gait generating controller parameters that are explored in a limited number of experiments. See Extension 2 for a comparison between the walking gaits on paper and PDMS-2, and Extension 5 for the details of the independent learning runs for all the surface adhesion experiments.

Table 3. Comparison of the learning performances for changing surface adhesion.

Test Surface	Type	Performance (mm)	
		Median	IQR
Paper	Standard GP-BO	0.00	1.59
	Mean transfer	1.60	0.46
PS	Standard GP-BO	0.00	0.00
	Mean transfer	0.00	0.61
PDMS-0	Standard GP-BO	0.00	0.00
	Mean transfer	1.60	0.75
PDMS-1	Standard GP-BO	0.00	0.00
	Mean transfer	0.81	1.45
PDMS-2	Standard GP-BO	0.00	0.67
	Mean transfer	1.07	0.46

4.4.2. Surface roughness. Next, we test the same robot on five surfaces with increasing roughness properties that are reported in Figure 2(b). The walking gait performances achieved during the learning runs for the two extreme surfaces, P800-grit ($R^q = 7.488\mu$ m) and P60-grit ($R^q = 77.195\mu$ m) sandpaper replica are shown in Figure 8(a)–(d). Similar to the surface adhesion experiments, the mean transfer method improves the learning performance of the standard BO by increasing the number of explored parameter sets that generate non-zero walking gait performances. The comparative performances of the standard BO and the mean transfer method for all the roughness test surfaces are reported in Figure 8(e) and Table 4. See Extension 3 for a comparison between the walking gaits on P800 and P60, and Extension 6 for the details of the independent learning runs for all the surface roughness experiments.

4.4.3. Medium viscosity. Finally, we test our robot walking in eight different media with changing viscosity as reported in Figure 2(c), and report the results for two extreme cases, cP1 and cP90 in Figure 9(a)–(d). It can be seen that the ability to explore a wide range of controller parameter sets that generate walking gaits decreases for both BO approaches for the high-viscosity fluids (cP $>$ 35). However, the mean transfer method still manages to increase the number of successful sets compared with the

Table 4. Comparison of the learning performances for changing roughness values.

Test Surface	Type	Performance (mm)	
		Median	IQR
P800	Standard GP-BO	0.00	1.11
	Mean transfer	1.18	0.59
P400	Standard GP-BO	0.48	1.06
	Mean transfer	1.07	0.79
P240	Standard GP-BO	0.00	0.00
	Mean transfer	1.14	0.35
P120	Standard GP-BO	0.00	0.82
	Mean transfer	0.37	1.14
P60	Standard GP-BO	0.00	0.00
	Mean transfer	0.00	1.01

Table 5. Comparison of the learning performances for changing medium viscosities.

Test Medium	Type	Performance (mm)	
		Median	IQR
cP1	Standard GP-BO	0.00	1.04
	Mean transfer	0.95	1.21
cP3	Standard GP-BO	0.00	1.25
	Mean transfer	1.25	1.44
cP6	Standard GP-BO	0.98	1.35
	Mean transfer	1.21	1.48
cP9	Standard GP-BO	0.00	1.32
	Mean transfer	1.21	0.70
cP20	Standard GP-BO	0.00	0.19
	Mean transfer	0.67	1.10
cP35	Standard GP-BO	0.00	0.00
	Mean transfer	0.00	0.99
cP60	Standard GP-BO	0.00	0.00
	Mean transfer	0.00	0.00
cP90	Standard GP-BO	0.00	0.00
	Mean transfer	0.00	0.00

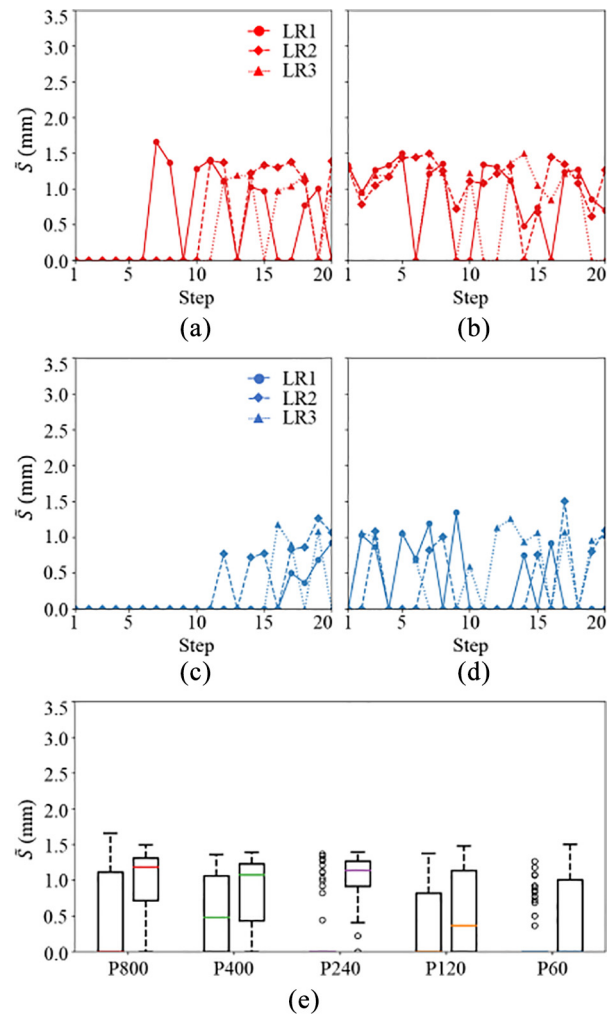

Fig. 8. The learning of the controller parameters for the changing roughness values of the test surfaces (upper row P800, lower row P60) within 20 physical experiments in 3 independent learning runs (a) and (c) without utilizing the prior information and (b) and (d) with utilizing the prior information. Overall performances of the learning runs with the standard BO (left bars) and the mean transfer method (right bars) reported with box plots for all test cases (e).

standard BO for all the media. The overall performances of the standard and mean transfer approaches are reported as box plots in Figure 9(e) and Table 5. These results are consistent with the other two test surfaces that the TL method allows the BO to explore more of the controller sets that generate walking gaits against the changing task space properties. See Extension 4 for a comparison between the walking inside cP1 and cP90, and Extension 7 for the details of the independent learning runs for all the medium viscosity experiments.

5. Discussion

The displacement measurements from the exhaustive search experiments in Figures 3(a)–(c) show that even though three identical robots are tested with the same controller parameters on the same surface, they generate different walking gait performances. These initial results also confirm the observations related to performance repeatability in our previous work (Culha et al., 2020). Moreover, the influence of the task space on the robot performance can be seen clearly from Figures 3(h)–(j), where the adhesion and roughness differences between different surfaces are reflected. These observations support the necessity of a data-efficient controller learning system that is robust to the robot performance variabilities caused by the material, fabrication, and the task environment of the small-scale, medical-oriented, and untethered soft robots.

Our choice on the application of BO and GPs to directly learn the controller parameters of our soft millirobot is based on three aspects. First, BO offers the efficient data-driven optimization of continuum and complex black-box

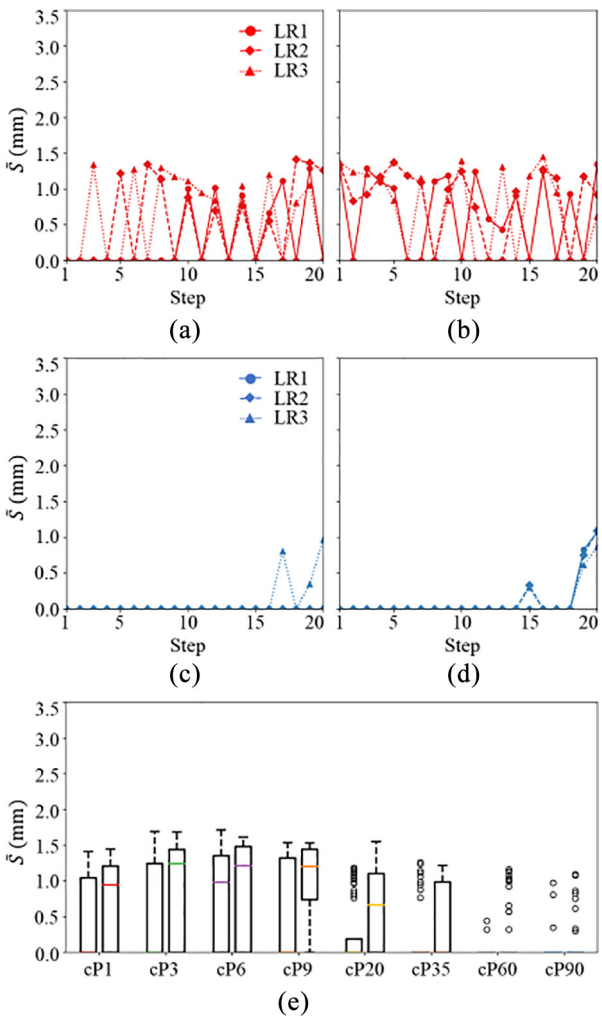

Fig. 9. The learning of the controller parameters for the changing test medium viscosity (upper row cP1, lower row cP90) within 20 physical experiments in 3 independent learning runs (a) and (c) without the prior information and (b) and (d) with the prior information. Overall performances of the learning runs with the standard BO (left bars) and the mean transfer approach (right bars) reported with box plots for all test cases (e).

functions that do not have a closed-form definition. This feature addresses our challenges with not having a deterministic model for the kinematics for our robot and requiring to achieve optimized walking gaits in a small number of experiments. Second, the investigated function can be represented with GPs within BO, which allows capturing the noise and unknown disturbances. As our robot inherits fabrication-, material-, and interaction-based performance disturbances, GPs provide us with a walking gait function representation that includes these variances. Finally, BO is a global optimization tool that avoids getting stuck at local minima, which is important for exploring the parameter space of the investigated function. Combined with the appropriate TL method, this feature allowed us to overcome local minima while adapting the controller parameters to different robots and task spaces. While we do not claim that BO is the best or only suitable optimization technique here, these three aspects successfully address the controller and modeling challenges existing for our robotic system and make the BO the choice of our application.

The experimental results in this study show that our approach of using BO with GPs and TL methods allowed a data-efficient (i.e., using as few experiments as possible) controller learning that achieves adaptation to different task spaces within a wide range (i.e., on the scale of an order of magnitude) of surface and medium properties. Our main goal is to allow the learning system to explore the controller parameter space to find more of the parameter sets that generate successful walking gaits in response to changing task environments. For this purpose, we configured our BO to favor exploration more than exploitation. That is why we do not focus on finding the optimum walking gait controller parameters for each robot or task space in our experiments. Consequently, our current approach does not establish a straightforward correlation between the change of controller parameters with respect to changing robot and task conditions. The comparative results between the standard BO and the TL methods show that both approaches can find sub-optimum parameter sets owing to the statistical nature of the learning method, whose results are given in Appendix C for the task space adaptation experiments. However, we propose that TL methods may allow the system to explore a larger portion of the function space in a fewer number of physical experiments, hence achieving data-efficiency in learning.

In terms of experimental learning efficiency, the transfer of the prior mean information outperformed the other TL methods in our experiments. The transfer of this information allowed the BO to start the parameter exploration in the function space within the regions of high-performance result expectations. Therefore, it took the BO much faster to explore the parameter spaces that generate optimum walking gaits (see Extension 8 for a sample comparison of parameter selection with the standard BO and mean transfer method). We see the same effect for the test case of robot 3 walking on P800 in Figure 5 (s). Here, this TL method allowed the exploration of the regions with higher expected results and surpassed the exploration boundary of the standard BO. The larger variance in the stride length performances explored by this TL method is caused by this exploration tendency. In comparison, we see that the HP-4 method failed to explore the controller parameters that yield optimum gaits because of the transfer of the signal variance parameter σ_f^2. When this parameter was optimized σ_f^2 for a single robot (i.e., robot 3 on paper in our case), it resulted in a smaller value than that used in standard BO and eventually it decreased the exploration weight in (11). Thus, the BO that started with this transferred parameter value focused more on the exploitation of known regions as soon as it found a parameter set

generating non-zero performance, and avoided exploring the unobserved regions for the remaining iteration steps in the learning runs. We see the influence of the hindered exploration in the first column of Figure 5. We see that the transfer of the length scale hyperparameters in HP-l_c was typically ineffective in our test cases because the length scale is bound to the robot geometry and magnetic profile, and we used robots with the same geometries in the experiments. The hybrid transfer approach, which also included the transfer of length scale hyperparameters, showed similar performances with the prior mean transfer. These similar results also show the ineffectiveness of length scales hyperparameters and the dominance of prior mean information. Although the hybrid approach could be extended by also transferring the variance terms σ_n^2 and σ_f^2 along with the prior mean information, we did not include these alternatives in the TL comparison experiments for two reasons. First, the negative effect of the signal variance σ_f^2 on the exploration capability was already shown and, second, σ_n^2, i.e., the noise in the collected dataset remained the same for all test cases as we used the same hardware setup for all our experiments. The values of the hyperparameters can also be seen in Appendix C. The choice of the parameters such as the kernels (Wilson, 2014) and hyperparameters (Chen and Wang, 2018) can also be replaced with other methods, however, the systematic analysis of their influence on learning performance is beyond the scope of our current study.

In the learning experiments that compare the standard BO and four TL approaches, we chose to represent learning performance with median and IQR instead of mean and standard deviation (as seen in Figure 4), since IQR is a robust measure of scale, as it is less sensitive to the outliers in the data. Moreover, dissimilar to standard deviation, IQR can represent the skewness in the distribution of the walking performance results, which becomes more apparent as the performance values get closer to the ends of possible performance ranges. In addition to its advantages in the statistical distribution representation, IQR does not report any unachievable result according to the gait definition in Section 2.2.

6. Conclusion and future work

In this study, we have investigated the use of BO with GPs to experimentally learn the controller parameters for the walking gait of a magnetic soft millirobot. We have created benchmark datasets consisting of 750 experimental results using an exhaustive search to find the walking gait function space for five different test cases. We then used these datasets to compare the effectiveness of four different TL methods to complement the standard BO learning. In these experiments that involve 104 learning steps for each test case, we have shown that the transfer of the prior mean information increased the BO learning performance the most in terms of increasing the number of explored sub-optimum controller parameters and decreasing the number of required experiments. Based on these findings, we also applied BO learning together with the prior mean transfer method on different task spaces with changing surface adhesion, surface roughness, and medium viscosity. We have shown that controller learning with a BO that utilizes prior mean transfer demonstrates successful adaptation to task spaces in a data-efficient way by exploring the function space of the robot in fewer experiments to find a larger group of controller parameters that yield successful walking gaits.

Our approach is not only limited to walking gait learning and it can further be applied to different locomotion and manipulation controllers for soft robots (Chin et al., 2020). In the future, studies focusing on small-scale fabrication with higher magnetization resolution may address the fabrication reproducibility issues (Alapan et al., 2020; Kim et al., 2018; Xu et al., 2019). However, especially for robots designed for biomedical operations, the interaction with the dynamic task environment may still have degrading robot material and performance effects. For such scenarios, a data-efficient controller learning system may adapt optimum controller parameters to these changes in the robot. For example, such an approach may be applied to endoscopic soft robots within or outside the gastrointestinal (GI) tract (Son et al., 2020; Yim et al., 2014) using a small number of trials. Our study can be further extended to involve the design parameters, such as the magnetic particle density in our robots, and guide the task-oriented design strategies for future soft mobile robots. Our approach can be used to reveal design guidelines to improve the kinematic models of the small-scale robots while utilizing the CC approximations (Webster and Jones, 2010), analytical models (Renda et al., 2014), and FEMs (Largilliere et al., 2015). However, as the BO we are using is an episodic algorithm, meaning that each suggested parameter set must be evaluated first in an experiment, the adaptation to design optimization will require the experiments to be run either in a simulation environment or an automated rapid fabrication system that can be integrated within the actuation architecture. The systematic comparison of our experimental approach to alternative optimization and control methods supported with simulations such as intelligent trial and error (Cully et al., 2015), evolution algorithms (Kriegman et al., 2020), or policy gradients (Sehnke et al., 2010) is beyond the scope of our current study but is an interesting task for future work. We believe that the benchmark datasets available in this study can be used to compare these different methods. Our long-term vision is to build fully autonomous systems that can

control, track, evaluate, and optimize soft robots operating in changing complex real-world environments, with minimum human involvement.

Acknowledgments

S.O.D. thanks the Ministry of National Education of the Republic of Turkey for the Doctoral Scholarship. U.C. and A.P.F. thank the Alexander von Humboldt Foundation for the Humboldt Postdoctoral Research Fellowship and the Federal Ministry for Education and Research.

Data accessibility statement

The experimental data from the benchmark datasets are available for access in https://github.com/sozgundemir/softrobotwalking datasethere.

Funding

This work was funded in part by the Alexander von Humboldt Foundation, the Ministry of National Education of the Republic of Turkey, the Cyber Valley Initiative, the Grassroots Initiative of the Max Planck Institute for Intelligent Systems, the Max Planck Society, and the European Research Council (ERC) Advanced Grant "SoMMoR" Project (grant number 834531).

ORCID iDs

Sinan Ozgun Demir https://orcid.org/0000-0002-6177-8372
Utku Culha https://orcid.org/0000-0002-7788-6124
Alp Karacakol https://orcid.org/0000-0003-1241-0800
Metin Sitti https://orcid.org/0000-0001-8249-3854

References

Alapan Y, Karacakol AC, Guzelhan SN, Isik I and Sitti M (2020) Reprogrammable shape morphing of magnetic soft machines. *Science Advances* 6(38): eabc6414.

Alexander RM (1984) The gaits of bipedal and quadrupedal animals. *The International Journal of Robotics Research* 3(2): 49–59.

Bern JM, Schnider Y, Banzet P, Kumar N and Coros S (2020) Soft robot control with a learned differentiable model. In: *2020 3rd IEEE International Conference on Soft Robotics (RoboSoft)*, pp. 417–423.

Brochu E, Cora VM and de Freitas N (2010) A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. *Preprint arXiv:1612.06830 [cs/LG]*.

Calandra R, Seyfarth A, Peters J and Deisenroth MP (2016) Bayesian optimization for learning gaits under uncertainty. *Annals of Mathematics and Artificial Intelligence* 76(1–2): 5–23.

Calisti M, Picardi G and Laschi C (2017) Fundamentals of soft robot locomotion. *Journal of The Royal Society Interface* 14(130): 20170101.

Chatzilygeroudis K, Vassiliades V, Stulp F, Calinon S and Mouret JB (2019) A survey on policy search algorithms for learning robot controllers in a handful of trials. *IEEE Transactions on Robotics* 36(2): 328–347.

Chen Z and Wang B (2018) How priors of initial hyperparameters affect Gaussian process regression models. *Neurocomputing* 275: 1702–1710.

Chenevier J, González D, Aguado JV, Chinesta F and Cueto E (2018) Reduced-order modeling of soft robots. *PloS One* 13(2): e0192052.

Chin K, Hellebrekers T and Majidi C (2020) Machine learning for soft robotic sensing and control. *Advanced Intelligent Systems* 2(6): 1900171.

Cianchetti M, Laschi C, Menciassi A and Dario P (2018) Biomedical applications of soft robotics. *Nature Reviews Materials* 3(6): 143–153.

Culha U, Demir SO, Trimpe S and Sitti M (2020) Learning of sub-optimal gait controllers for magnetic walking soft millirobots. In: *Proceedings of Robotics: Science and Systems (RSS)*.

Cully A, Clune J, Tarapore D and Mouret JB (2015) Robots that can adapt like animals. *Nature* 521(7553): 503–507.

Delhaisse B, Esteban D, Rozo L and Caldwell D (2017) Transfer learning of shared latent spaces between robots with similar kinematic structure. In: *2017 International Joint Conference on Neural Networks (IJCNN)*, pp. 4142–4149.

Della Santina C, Bicchi A and Rus D (2020) On an improved state parametrization for soft robots with piecewise constant curvature and its use in model based control. *IEEE Robotics and Automation Letters* 5(2): 1001–1008.

Devin C, Gupta A, Darrell T, Abbeel P and Levine S (2017) Learning modular neural network policies for multi-task and multi-robot transfer. In: *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2169–2176.

Duvenaud DK, Nickisch H and Rasmussen C (2011) Additive Gaussian processes. In: *Advances in Neural Information Processing Systems*, volume 24. Curran Associates, Inc.

Ghahramani Z (2015) Probabilistic machine learning and artificial intelligence. *Nature* 521(7553): 452–459.

Goury O and Duriez C (2018) Fast, generic, and reliable control and simulation of soft robots using model order reduction. *IEEE Transactions on Robotics* 34(6): 1565–1576.

GPy (2012) GPy: A Gaussian process framework in Python. Available at: http://github.com/SheffieldML/GPy.

Grazioso S, Di Gironimo G and Siciliano B (2019) A geometrically exact model for soft continuum robots: The finite element deformation space formulation. *Soft Robotics* 6(6): 790–811.

Gu H, Boehler Q, Cui H, et al. (2020) Magnetic cilia carpets with programmable metachronal waves. *Nature Communications* 11(1): 1–10.

Haddadin S, De Luca A and Albu-Schäffer A (2017) Robot collisions: A survey on detection, isolation, and identification. *IEEE Transactions on Robotics* 33(6): 1292–1312.

Helwa MK and Schoellig AP (2017) Multi-robot transfer learning: A dynamical system perspective. In: *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4702–4708.

Hiller J and Lipson H (2014) Dynamic simulation of soft multimaterial 3D-printed objects. *Soft Robotics* 1(1): 88–101.

Hu W, Lum GZ, Mastrangeli M and Sitti M (2018) Small-scale soft-bodied robot with multimodal locomotion. *Nature* 554(7690): 81–85.

Huang W, Huang X, Majidi C and Jawed MK (2020) Dynamic simulation of articulated soft robots. *Nature Communications* 11(2233): 1–9.

Hughes J, Culha U, Giardina F, Guenther F and Rosendo A Iida F (2016) Soft manipulators and grippers: A review. *Frontiers in Robotics and AI* 3: 69.

Hyatt P, Wingate D and Killpack MD (2019) Model-based control of soft actuators using learned non-linear discrete-time models. *Frontiers in Robotics and AI* 6(22): 22.

Iida F and Nurzaman SG (2016) Adaptation of sensor morphology: An integrative view of perception from biologically inspired robotics perspective. *Interface Focus* 6(4): 20160016.

Jeong SH, Zhang S, Hjort K, Hilborn J and Wu Z (2016) PDMS-based elastomer tuned soft, stretchable, and sticky for epidermal electronics. *Advanced Materials* 28(28): 5830–5836.

Jones DR, Schonlau M and Welch WJ (1998) Efficient global optimization of expensive black-box functions. *Journal of Global Optimization* 13(4): 455–492.

Kim Y, Parada GA, Liu S and Zhao X (2019) Ferromagnetic soft continuum robots. *Science Robotics* 4(33): aax7329.

Kim Y, Yuk H, Zhao R, Chester SA and Zhao X (2018) Printing ferromagnetic domains for untethered fast-transforming soft materials. *Nature* 558(7709): 274–279.

Kriegman S, Blackiston D, Levin M and Bongard J (2020) A scalable pipeline for designing reconfigurable organisms. *Proceedings of the National Academy of Sciences* 117(4): 1853–1859.

Largilliere F, Verona V, Coevoet E, Sanz-Lopez M, Dequidt J and Duriez C (2015) Real-time control of soft-robots using asynchronous finite element modeling. In: *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2550–2555.

Laschi C, Mazzolai B and Cianchetti M (2016) Soft robotics: Technologies and systems pushing the boundaries of robot abilities. *Science Robotics* 1(1): eaah3690.

Liao T, Wang G, Yang B, et al. (2019) Data-efficient learning of morphology and controller for a microrobot. In: *2019 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2488–2494.

Lu H, Zhang M, Yang Y, et al. (2018) A bioinspired multilegged soft millirobot that functions in both dry and wet conditions. *Nature Communications* 9(1): 1–7.

Majidi C (2014) Soft robotics: A perspective - current trends and prospects for the future. *Soft Robotics* 1(1): 5–11.

Makondo N, Rosman B and Hasegawa O (2018) Accelerating model learning with inter-robot knowledge transfer. In: *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2417–2424.

Marco A, Baumann D, Khadiv M, Hennig P, Righetti L and Trimpe S (2020) Robot learning with crash constraints. *arXiv Preprint arXiv:2010.08669*.

Pan SJ and Yang Q (2009) A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering* 22(10): 1345–1359.

Patacchiola M, Turner J, Crowley EJ, O'Boyle M and Storkey A (2020) Bayesian meta-learning for the few-shot setting via deep kernels. In: *Advances in Neural Information Processing Systems (NeurIPS)*.

Perrone V, Shen H, Seeger MW, Archambeau C and Jenatton R (2019) Learning search spaces for Bayesian optimization: Another view of hyperparameter transfer learning. In: *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 12771–12781.

Polygerinos P, Correll N, Morin SA, et al. (2017) Soft robotics: Review of fluid-driven intrinsically soft devices; manufacturing, sensing, control, and applications in human–robot interaction. *Advanced Engineering Materials* 19(12): 1700016.

Raina R, Ng AY and Koller D (2006) Constructing informative priors using transfer learning. In: *Proceedings of the 23rd International Conference on Machine Learning (ICML)*, pp. 713–720.

Rasmussen CE and Williams CKI (2006) *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. Cambridge, MA: The MIT Press.

Ren Z, Hu W, Dong X and Sitti M (2019) Multi-functional soft-bodied jellyfish-like swimming. *Nature Communications* 10(1): 1–12.

Renda F, Boyer F, Dias J and Seneviratne L (2018) Discrete cosserat approach for multisection soft manipulator dynamics. *IEEE Transactions on Robotics* 34(6): 1518–1533.

Renda F, Giorelli M, Calisti M, Cianchetti M and Laschi C (2014) Dynamic model of a multibending soft robot arm driven by cables. *IEEE Transactions on Robotics* 30(5): 1109–1122.

Rich SI, Wood RJ and Majidi C (2018) Untethered soft robotics. *Nature Electronics* 1(2): 102.

Rieffel J and Mouret JB (2018) Adaptive and resilient soft tensegrity robots. *Soft Robotics* 5(3): 318–329.

Rus D and Tolley MT (2015) Design, fabrication and control of soft robots. *Nature* 521(7553): 467–475.

Schramm L, Sintov A and Boularias A (2020) Learning to transfer dynamic models of underactuated soft robotic hands. In: *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4579–4585.

Sehnke F, Osendorfer C, Rückstieß T, Graves A, Peters J and Schmidhuber J (2010) Parameter-exploring policy gradients. *Neural Networks* 23(4): 551–559.

Shahriari B, Swersky K, Wang Z, Adams RP and De Freitas N (2015) Taking the human out of the loop: A review of Bayesian optimization. *Proceedings of the IEEE* 104(1): 148–175.

Shen Z, Chen F, Zhu X, Yong KT and Gu G (2020) Stimuli-responsive functional materials for soft robotics. *Journal of Materials Chemistry B* 8(39): 8972–8991.

Sitti M (2018) Miniature soft robots - road to the clinic. *Nature Reviews Materials* 3(6): 74–75.

Son D, Gilbert H and Sitti M (2020) Magnetically actuated soft capsule endoscope for fine-needle biopsy. *Soft Robotics* 7(1): 10–21.

Thuruthel TG, Ansari Y, Falotico E and Laschi C (2018) Control strategies for soft robotic manipulators: A survey. *Soft Robotics* 5(2): 149–163.

Thuruthel TG, Shih B, Laschi C and Tolley MT (2019) Soft robot perception using embedded soft sensors and recurrent neural networks. *Science Robotics* 4(26): aav1488.

Trimmer BA and Lin HT (2014) Bone-free: Soft mechanics for adaptive locomotion. *Integrative and Comparative Biology* 54(6): 1122–1135.

von Rohr A, Trimpe S, Marco A, Fischer P and Palagi S (2018) Gait learning for soft microrobots controlled by light fields. In: *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 6199–6206.

Walsh C (2018) Human-in-the-loop development of soft wearable robots. *Nature Reviews Materials* 3(6): 78–80.

Wang K, Chen J, Xie L and Su H (2020) Transfer learning based on incorporating source knowledge using Gaussian process

models for quick modeling of dynamic target processes. *Chemometrics and Intelligent Laboratory Systems* 198: 103911.

Webster RJ III and Jones BA (2010) Design and kinematic modeling of constant curvature continuum robots: A review. *The International Journal of Robotics Research* 29(13): 1661–1683.

Wilson AG (2014) Covariance kernels for fast automatic pattern discovery and extrapolation with Gaussian processes. PhD Thesis, University of Cambridge Cambridge, UK.

Wu Y, Yim JK, Liang J, et al. (2019) Insect-scale fast moving and ultrarobust soft robot. *Science Robotics* 4(32): eaax1594.

Xu T, Zhang J, Salehizadeh M, Onaizah O and Diller E (2019) Millimeter-scale flexible robots with programmable three-dimensional magnetization and motions. *Science Robotics* 4(29): eaav4494.

Yang B, Wang G, Calandra R, Contreras D, Levine S and Pister K (2018) Learning flexible and reusable locomotion primitives for a microrobot. *IEEE Robotics and Automation Letters* 3(3): 1904–1911.

Yim S, Gultepe E, Gracias DH and Sitti M (2014) Biopsy using a magnetic capsule endoscope carrying, releasing, and retrieving untethered microgrippers. *IEEE Transactions on Biomedical Engineering* 61(2): 513–521.

Appendix A. Index to multimedia extensions

Archives of IJRR multimedia extensions published prior to 2014 can be found at http://www.ijrr.org, after 2014 all videos are available on the IJRR YouTube channel at http://www.youtube.com/user/ijrrmultimedia

Table of Multimedia Extensions.

Extension	Media type	Description
1	Video	Robot tracking and gait detection output.
2	Video	Comparison of walking on paper and PDMS-2.
3	Video	Comparison of walking gaits on P800 and P60.
4	Video	Comparison of walking gaits inside cP1 and cP90.
5	Image	Learning run details of the complete set of adhesive surfaces.
6	Image	Learning run details of the complete set of rough surfaces.
7	Image	Learning run details of the complete set of viscous media.
8	Video	Sample learning runs comparing the parameter selection process by a standard BO with prior mean transfer.

Appendix B. Pseudo-code for robot tracking and motion evaluation algorithm

The pseudo-code for robot tracking and motion evaluation is given in Algorithm 1.

Algorithm 1. Robot Tracking and Motion Evaluation.

```
while Test is running do
  Image ← Capture robot's image by orthogonal camera,
  Image ← Correct distortion and misalignment in the Image,
  RobotImage ← Apply threshold on Image to find robot pixels,
  RobotLegs ← Find leg positions running Grassfire
    algorithm on RobotImage,
  for Leg ∈ RobotLegs do
    LegState ← Label leg state as 'touchDown', 'liftOff', or
      'slipping',
    RobotState ← Label robot state using LegState
    MotionType ← Label robot's motion either as walking or
      not by comparing RobotState throughout the actuation
      sequence with the four states in Figure 1(f)
```

Appendix C. Hyperparameter and controller parameter sets

Tables 6–9 list the exact values of the GP hyperparameters used in the TL experiments and the controller parameter sets in the experiments for the adaptation to task spaces, and Table 10 lists the nomenclature used throughout the article.

Table 6. Hyperparameter sets used in the benchmark tests.

Learning Method	σ_n	σ_f	$l_c^{\alpha_0}$	$l_c^{\alpha_1}$
Standard	0.29	1.85	17.75	15.25
HP-4	0.29	0.80	10.12	9.99
HP-l_c	0.29	1.85	12.47	12.74
Mean	0.29	1.85	17.75	15.25
Hybrid	0.29	1.85	12.47	12.74

Table 7. Best-performing controller parameter sets and the corresponding stride length averages for the changing adhesion strength characteristics of the test surfaces within 20 physical experiments in 3 independent learning runs (60 experiments in total) with and without utilizing the prior information.

Surface	Type	Controller parameters				Stride length S (mm) (average ± SD)
		B_{max} (mT)	f (Hz)	α_1 (°)	α_2 (°)	
Paper	Standard	11.0	0.5	0.0	80.0	1.48 ± 0.15
	TL	12.0	0.5	0.0	90.0	1.29 ± 0.12
PS	Standard	8.0	0.5	35.0	65.0	0.94 ± 0.03
	TL	12.0	4.0	35.0	90.0	1.46 ± 0.05
PDMS-0	Standard	12.0	1.0	50.0	85.0	1.18 ± 0.05
	TL	11.0	0.5	10.0	80.0	1.49 ± 0.19
PDMS-1	Standard	5.0	0.5	0.0	90.0	0.56 ± 0.06
	TL	12.0	0.5	0.0	85.0	1.37 ± 0.34
PDMS-2	Standard	12.0	0.5	30.0	90.0	1.29 ± 0.06
	TL	12.0	0.5	20.0	90.0	1.24 ± 0.05

Table 8. Best-performing controller parameter sets and the corresponding stride length averages for the changing roughness values of the test surfaces within 20 physical experiments in 3 independent learning runs (60 experiments in total) with and without utilizing the prior information.

Surface	Type	Controller parameters				Stride length S (mm) (average ± SD)
		B_{max} (mT)	f (Hz)	α_1 (°)	α_2 (°)	
P800	Standard	12.0	0.5	0.0	90.0	1.28 ± 0.03
	TL	12.0	4.0	30.0	85.0	1.36 ± 0.05
P400	Standard	12.0	0.5	15.0	70.0	1.18 ± 0.11
	TL	12.0	0.5	40.0	90.0	1.25 ± 0.10
P240	Standard	12.0	0.5	20.0	85.0	1.34 ± 0.08
	TL	12.0	0.5	25.0	90.0	1.28 ± 0.04
P120	Standard	12.0	0.5	0.0	90.0	1.17 ± 0.11
	TL	12.0	2.0	30.0	80.0	1.27 ± 0.13
P60	Standard	12.0	4.0	30.0	85.0	1.26 ± 0.11
	TL	12.0	6.0	30.0	80.0	1.50 ± 0.11

Table 9. Best-performing controller parameter sets and the corresponding stride length averages for the changing viscosity values of the test medium within 20 physical experiments in 3 independent learning runs (60 experiments in total) with and without utilizing the prior information.

Medium	Type	Controller parameters				Stride length S (mm) (average ± SD)
		B_{max} (mT)	f (Hz)	α_1 (°)	α_2 (°)	
cP1	Standard	12.0	0.5	25.0	90.0	1.41 ± 0.03
	TL	11.0	4.0	25.0	75.0	1.45 ± 0.05
cP3	Standard	10.0	1.0	25.0	80.0	1.69 ± 0.04
	TL	12.0	0.5	0.0	80.0	1.69 ± 0.04
cP6	Standard	12.0	0.5	15.0	85.0	1.72 ± 0.09
	TL	12.0	0.5	10.0	90.0	1.61 ± 0.08
cP9	Standard	12.0	0.5	5.0	80.0	1.54 ± 0.02
	TL	12.0	0.5	30.0	75.0	1.53 ± 0.03
cP20	Standard	9.0	0.5	35.0	65.0	1.18 ± 0.05
	TL	9.0	0.5	0.0	85.0	1.55 ± 0.14
cP35	Standard	12.0	0.5	1.0	60.0	1.26 ± 0.10
	TL	11.0	0.5	25.0	65.0	1.22 ± 0.06
cP60	Standard	12.0	20.0	30.0	75.0	0.44 ± 0.14
	TL	12.0	0.5	25.0	65.0	1.16 ± 0.03
cP90	Standard	5.0	0.5	10.0	65.0	0.97 ± 0.16
	TL	8.0	0.5	30.0	65.0	1.09 ± 0.11

Table 10. Nomenclature.

B	Magnetic field
$B_{\max}$	Maximum magnetic field magnitude
f	Frequency of the actuation cycle
$\alpha_{1,2}$	Two magnetic field orientation angles
L	Robot length
w	Robot width
h	Robot height
R_q	Root mean square height
S	Stride length
θ	Controller parameter set
Θ	Complete search space for θ
$S(\theta)$	Average stride length performance of the robot for a given θ
n	Zero-mean Gaussian noise in each experimental measurement
σ_n^2	Variance of the noise n
$\tilde{S}$	Observed stride length measured with noise n
$\mu(\theta)$	Prior mean for a given set θ
k	Kernel (i.e., the covariance function) of the GP
D	Experimental data
N	Number of experiments
$\delta_{i,j}$	Kronecker delta for experiments i and j
l_c	Length scale
σ_f^2	Variance in the signal
α_{acq}	Acquisition function
Φ	Gaussian cumulative density function
ϕ	Gaussian probability density function
ξ	GP exploration/exploitation weight
S'	E xperimental displacement measurement
$\hat{S}$	N ormalized gait performance
ε	N ormalized error of the stride length performance