EXOSKELETONS

Deep domain adaptation eliminates costly data required for task-agnostic wearable robotic control

Keaton L. Scherpereel[1,2]*†, Matthew C. Gombolay[2,3], Max K. Shepherd[4], Carlos A. Carrasquillo[1,2], Omer T. Inan[5], Aaron J. Young[1,2]

Data-driven methods have transformed our ability to assess and respond to human movement with wearable robots, promising real-world rehabilitation and augmentation benefits. However, the proliferation of data-driven methods, with the associated demand for increased personalization and performance, requires vast quantities of high-quality, device-specific data. Procuring these data is often intractable because of resource and personnel costs. We propose a framework that overcomes data scarcity by leveraging simulated sensors from biomechanical models to form a stepping-stone domain through which easily accessible data can be translated into data-limited domains. We developed and optimized a deep domain adaptation network that replaces costly, device-specific, labeled data with open-source datasets and unlabeled exoskeleton data. Using our network, we trained a hip and knee joint moment estimator with performance comparable to a best-case model trained with a complete, device-specific dataset [incurring only an 11 to 20%, 0.019 to 0.028 newton-meters per kilogram (Nm/kg) increase in error for a semisupervised model and 20 to 44%, 0.033 to 0.062 Nm/kg for an unsupervised model]. Our network significantly outperformed counterpart networks without domain adaptation (which incurred errors of 36 to 45% semisupervised and 50 to 60% unsupervised). Deploying our models in the real-time control loop of a hip/knee exoskeleton ($N = 8$) demonstrated estimator performance similar to offline results while augmenting user performance based on those estimated moments (9.5 to 14.6% metabolic cost reductions compared with no exoskeleton). Our framework enables researchers to train real-time deployable deep learning, task-agnostic models with limited or no access to labeled, device-specific data.

INTRODUCTION

State-of-the-art exoskeleton control architectures increasingly use data-driven techniques to provide more effective assistance during human movement compared with their heuristic counterparts. In these approaches, data-driven models are trained to map data from exoskeleton-mounted sensors to human states [e.g., speed (*1*, *2*), gait phase (*3*–*5*), locomotion mode (*6*, *7*), kinematics (*8*, *9*), and joint moments (*10*–*12*)], environmental conditions [e.g., ground slope (*2*, *13*) or terrain (*14*, *15*)], or human outcomes [e.g., metabolic cost (*16*, *17*) or user preference (*18*)]. When sufficient labeled data (data paired with the correct output values) are available, many of these data-driven models, such as deep learning–based controllers, are capable of user independence (i.e., do not require participant-specific data) and can apply effective assistance across a wide range of tasks without requiring task classification (task agnostic) (*11*, *12*). These advances are pushing exoskeletons closer to handling the human and environmental uncertainty requisite for real-world viability.

Despite these successes, a critical barrier to the development and deployment of these data-driven controllers remains: A substantial amount of labeled training data is required, and these labels are costly to obtain for each device. Researchers must record data from their exoskeleton sensors along with time-synced data from a ground truth source, often necessitating optical motion capture systems and force plates. For instance, these motion and force data are critical to accurately calculate ground truth biological joint moments, which have recently emerged as a powerful human state variable in exoskeleton control (*11*, *12*). This highly specialized equipment is not readily available for many roboticists, and the process of calculating joint moments (or other human state labels) is nontrivial, requiring hundreds of hours of skilled labor for curation of large datasets. The training data are also device specific, prohibiting the reusage of data from previous generations or prototypes; every time researchers add or move sensors or substantially change the control type or magnitude, new labeled data must be collected. Furthermore, the exoskeleton must be actuated during training data collection because of the nonnegligible effect of assistance on users' biomechanics and sensor shifts due to soft tissue deformation and interface compliance. To accomplish this, researchers often must develop hand-tuned, task-dependent "stand-in" controllers that approximate the intended data-driven controller (*11*, *12*). Thus, progress in exoskeleton control research is substantially hindered by its dependence on device-specific, labeled data, which is only compounded by the short relevance time frame and numerous prohibitive barriers to obtaining that data.

Transfer learning is a set of techniques that have emerged to address problems like this by leveraging large, related datasets to more effectively learn to model smaller-scale, labeled data. For lower-limb devices, transfer learning methods have been applied to enable interparticipant and interactivity transfer of deep learning models. However, these techniques are limited in that some participant- or activity-specific data must exist and the specific sensors must remain the same (*19*–*21*). These studies have also been restricted to off-device verification. Researchers have also sought to create larger-scale, labeled, synthetic datasets for training deep learning networks, but this work has focused on dataset augmentation without

[1]George W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA. [2]Institute for Robotics and Intelligent Machines, Georgia Institute of Technology, Atlanta, GA 30332, USA. [3]School of Interactive Computing, Georgia Institute of Technology, Atlanta, GA 30332, USA. [4]College of Engineering, Bouvé College of Health Sciences, and Institute for Experiential Robotics, Northeastern University, Boston, MA 02115, USA. [5]School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA.
*Corresponding author. Email: keaton@gatech.edu
†Present address: Skip Innovations, San Francisco, CA, USA.

directly addressing the simulation-to-real gap present when deploying models trained with synthetic data in the real world (*22*, *23*). Thus, the feasibility of integrating large open-source datasets directly into real-time deployable algorithms for exoskeleton control remains an unanswered question.

Domain adaptation, a type of transfer learning, leverages labeled data from a related source domain to improve model performance in a target domain in which few or no labels exist. Domain adaptation techniques strive to uncover common latent features between the target and source distributions, enabling a mapping from unlabeled examples in the target domain to example-label pairs from the source domain. Relevant work from this area has shown that domain adaptation techniques can handle variability in wearable sensor location (*24*–*27*), sensor type (*28*), and user usage (*26*). However, these studies seldom go beyond simple classification tasks toward the important task of translating raw time series data (*26*, *29*), and we are unaware of any prior work that has leveraged domain adaptation in wearable robotics.

To enable scalable deep learning for exoskeletons, we created a domain adaptation framework that overcomes the need for difficult and expensive data collection by leveraging the following two simple-to-obtain data sources: open-source biomechanics datasets, which contain ground truth biological joint moment labels but no exoskeleton (termed the source domain), and data collected from the exoskeleton while participants perform a variety of tasks but without these moment labels (unlabeled data in the target domain). For example, when training end-to-end exoskeleton controllers, such as controllers built off an instantaneous estimate of biological joint moment (*11*, *12*), the most valuable data are also the most expensive to collect: exoskeleton sensor data with time-synced joint moment labels (i.e., labeled data in the target domain). By creating this domain adaptation framework, we can leverage unlabeled exoskeleton data, easily collected outside of the laboratory, to reduce or even eliminate the need for expensive data collection. This would greatly increase the ability to deploy exoskeletons at scale.

The key to our approach is a domain adaptation framework that leverages generative adversarial networks (GANs) formed into a CycleGAN (cycle-consistent generative adversarial network) and built on U-Net backbones to bridge the source and target domains. This bridge is formed by using a stepping-stone domain that simulates sensor data for any sensor location on the basis of samples in open-source biomechanics datasets and thus serves as a common, shared domain across devices and datasets [OpenSim (*30*, *31*)]. This sensor domain is readily developed on the basis of rigid body dynamics and serves as an intermediate representation that allows for the domain adaptation from real, open-source, biomechanics datasets to our target domain. Our framework translates human movement data between this simulated sensor source domain and the real sensor target domain (Fig. 1), enabling the integration of preexisting, large open-source biomechanics datasets in end-to-end exoskeleton controllers. This translation can be performed without any labeled data in the target domain (i.e., unsupervised), or we can supplement the GAN framework with a small-scale, labeled dataset in the target domain to ground the learned representation (i.e., semisupervised). Thus, we can achieve the best of both worlds: accurate predictions of biological moments while wearing an exoskeleton without the need to collect a large-scale dataset across all users and tasks.

In this study, we measured the usefulness of our translation by using the translated data to train downstream deep learning models in the real domain, which can be validated by deploying those models in real time on an exoskeleton. We compared end-to-end joint moment estimation models trained with a minimal number of labeled samples (semisupervised) and no labeled samples (unsupervised) to their respective equivalent baseline models without domain adaptation. These baselines are the best model that could be created from the limited data available using traditional supervised learning techniques and thus represent the current state-of-the-art approach. Comparing with these baselines represents the benefit that our domain adaption approach is providing in comparison with the current standard approach. This work represents three primary contributions: First, we proposed the use of human biomechanical models with simulated sensors as a universal domain for aggregating data with and without a device. Second, we developed a network that translates between this universal source domain and the device-specific, real-sensor, target domain. Last, we validated our translation strategy by deploying real-time models trained from translated

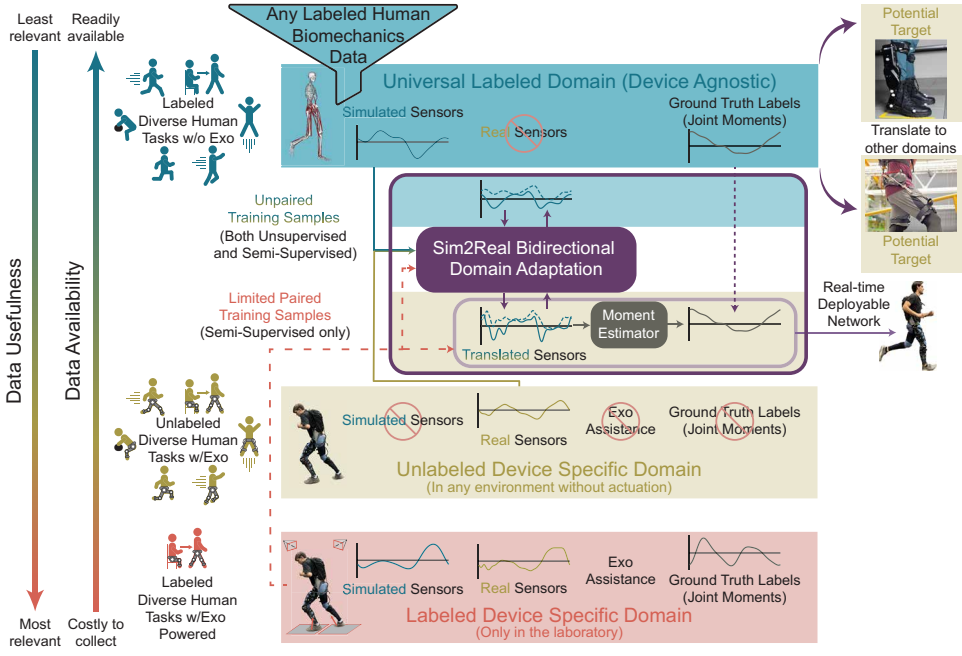

Fig. 1. Our strategy for replacing costly device-specific data with less costly data. Our approach uses the biomechanical modeling domain with simulated sensors as a universal domain for aggregating data. To use these data, we propose a network that performs bidirectional domain adaptation to translate simulated sensors into any specific device domain based on unlabeled data from that device. These translated data can be used to train downstream deep learning models (for our case, a moment estimator) that are deployable in real time on a device. This strategy can be used to create models for new devices and new joints. A fully unlabeled framework (unsupervised) is possible, but the optional labeled device-specific data can be useful (semisupervised).

data on an exoskeleton. These contributions allow our data-limited models to statistically outperform their counterpart models without domain adaptation (reducing error by ~15 to 30%) while also achieving accuracies similar (within ~10 to 30%) to those of models trained with large, costly datasets (best-case models). Accuracies like these allow exoskeleton controllers based on our model estimates to augment human performance, reducing metabolic expenditure by up to 10 to 15% on the basis of the two tasks captured. These are vital and critical steps toward enabling universal access to data-driven approaches in wearable robotics.

RESULTS

Optimization results

To train our semisupervised model, we needed to select a minimal subset of data for the supervised portion. Given that not all datasets are equally valuable, we sought to optimize the number of tasks (e.g., running, jumping, etc.) included in the training set by sequentially adding labeled tasks on the basis of the average generalization score: a score measuring how well each additional task improved moment estimation on the other tasks (Fig. 2A). Figure 2B shows the corresponding root mean square error (RMSE) across all tasks averaged between hip and knee for each task addition. Even after additional tasks no longer improve generalization, overall RMSE can continue to decrease because of improvement on that particular added task. We then compared the performance of our approach with that of a baseline model with the same number of tasks but without domain adaptation. Performance is presented as a percentage error increase compared with a best-case model (i.e., a model with access to all labeled tasks) for both the hip and knee (Fig. 2, C and D). Using the optimized set of tasks for both our approach and the baseline, our approach statistically outperformed the baseline approach regardless of the number of tasks included.

Although additional tasks continue to improve estimation, the goal of this optimization was to select a small, labeled subset of data to use from the target domain. Thus, we chose to limit our semisupervised approach to four tasks based on the plateau in the generalization score (Fig. 2A). The resulting percentage error increase from the best-case model at four tasks was 12.0 ± 3.8% at the hip and 13.6 ± 6.9% at the knee. For all future semisupervised analyses, the baseline and our approach included labeled exoskeleton data from only four optimized tasks. For our approach, these tasks were calisthenics, jump in place, level ground walk, and twister, whereas the baseline tasks were level ground walk, standing poses, calisthenics, and push-and-pull recovery (*12*).

Collecting many tasks is time and labor intensive, as is collecting additional participants; thus, we also sought to limit the number of participants included in our small semisupervised dataset. By folding across left-out participants, we assessed the average performance benefit of including additional participants. The performances of the baseline approach versus our approach for the hip and knee are presented in Fig. 2 (E and F), in terms of percentage increase compared with the same best-case model as above (all participants and tasks included). Only the four optimized tasks from above were included for each participant. Regardless of the number of participants included, our approach statistically outperformed the baseline approach. Given that most of the performance benefits were accrued during the first several participant additions, we limited the number of labeled participants who were included for the semisupervised models. We did this by averaging the hip and knee RMSE for our approach and determining the number of participants needed to reach 5% of the RMSE with all 13 participants. This occurred at four participants with a percentage increase in RMSE compared with the best case of 13.7 ± 5.2% at the hip and 19.7 ± 9.7% at the knee. Thus, on the basis of these optimizations, the following semisupervised models (both our approach and the baseline) included labeled data from only four participants performing four tasks.

Performance on the offline testing set

With the small, labeled dataset chosen for the semisupervised case, we tested the following approaches on a set of eight novel, held-out participants performing 28 tasks: the semisupervised and unsupervised baseline (i.e., without domain adaptation), our semisupervised and unsupervised approach, and the best-case model (i.e., using costly device-specific data from all participants and all tasks). For the semisupervised case, our approach significantly outperformed the baseline approach at the hip (percentage increase in RMSE compared with the best case of 11.3% versus the baseline of 35.5%, $P < 0.01$) and at the knee (percentage increase in RMSE compared with the best case of 20.3% versus the baseline of 45.3%, $P < 0.01$) as seen in Fig. 3 (A and B). For the unsupervised case, our approach significantly outperformed the baseline approach at the hip (percentage increase in RMSE compared with the best case of 19.9% versus the baseline of 49.5%, $P < 0.01$) and at the knee (percentage increase in RMSE compared with the best case of 44.4% versus the baseline of 60.3%, $P < 0.01$) as seen in Fig. 3 (C and D). Figure 3 (E and F) presents a comparison between the baseline and our approach for both the semisupervised and unsupervised cases in terms of the square of the Pearson correlation coefficient [R^2; on the best fit line between the ground truth and the estimate (*11*, *12*)]. Each point on the scatter plot represents one of the 28 tasks where a higher R^2 at the hip and the knee (closer to the top right-hand corner) indicates a better shape match between the estimate and the ground truth joint moment.

To further validate that our translation approach works across different types of estimators, we used our bidirectional translation approach to train a ground reaction force (GRF) estimator model (for both the supervised and unsupervised cases) and an activity classifier (for the unsupervised case). Both models demonstrated improved performance with our approach as compared with the baseline (figs. S4 to S6).

Real-time model performance

We deployed the joint moment estimation models trained with both our approach and the baseline in real time on an autonomous hip/knee exoskeleton with the models providing estimated joint moments to assist users (Fig. 4, A and B). The best-case model estimate was computed post hoc on the collected data for a fair comparison and performed similarly to the current state-of-the-art estimators (fig. S7) (*12*). Across eight novel users performing eight representative tasks, the model trained with our semisupervised approach significantly outperformed the baseline approach at the hip (percentage increase in RMSE compared with the best-case model of 11.0 ± 6.4% versus the baseline of 37.5 ± 10.5%, $P < 0.01$) and at the knee (percentage increase in RMSE compared with the best-case model of 19.5 ± 9.6% versus the baseline of 48.3 ± 19.0%, $P = 0.011$) (Fig. 4, C and D). The final performance of our semisupervised approach was 0.19 ± 0.01 Nm/kg RMSE and 0.71 ± 0.04 R^2 at the hip and 0.16 ± 0.02 Nm/kg RMSE and 0.75 ± 0.04 R^2 at the knee. For the

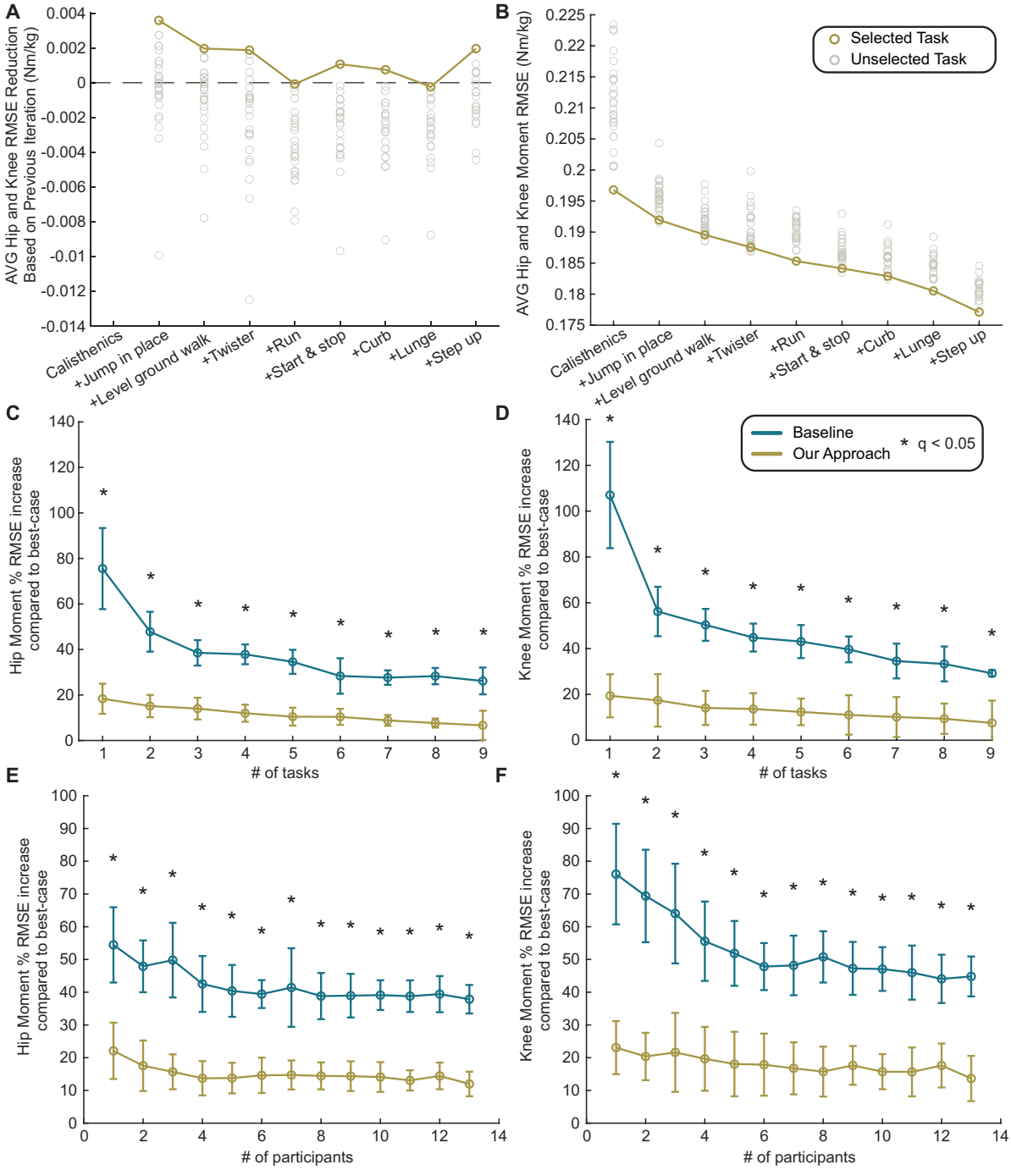

Fig. 2. Task optimization and performance across limited task and participant training sets. After seeding the task optimization with the task that provided the lowest RMSE across all tasks, tasks were sequentially added by maximizing the generalization score to other tasks (**A**). The corresponding RMSE for unselected and selected tasks is shown in (**B**). Optimization was terminated after nine tasks when our approach was not statistically different from the best-case model. Using this optimized task list, we compared the performance of our approach at the hip (**C**) and knee (**D**) with a separately optimized baseline without domain adaptation (but the same limited number of tasks). By cross-folding across participants with four optimized tasks, we assessed the average RMSE for our approach and the baseline for the hip (**E**) and knee (**F**) when sequentially increasing the number of labeled participants used in training. Our approach and the baseline were compared with a best-case model trained with labeled data from all tasks and participants. Error bars represent SD across five cross-folded, left-out participants [omitted from (A) and (B) for clarity]. Statistical significance was determined by controlling the false rate of discovery ($q < 0.05$) across all nine tasks [(C) and (D)] and all 13 participants [(E) and (F)].

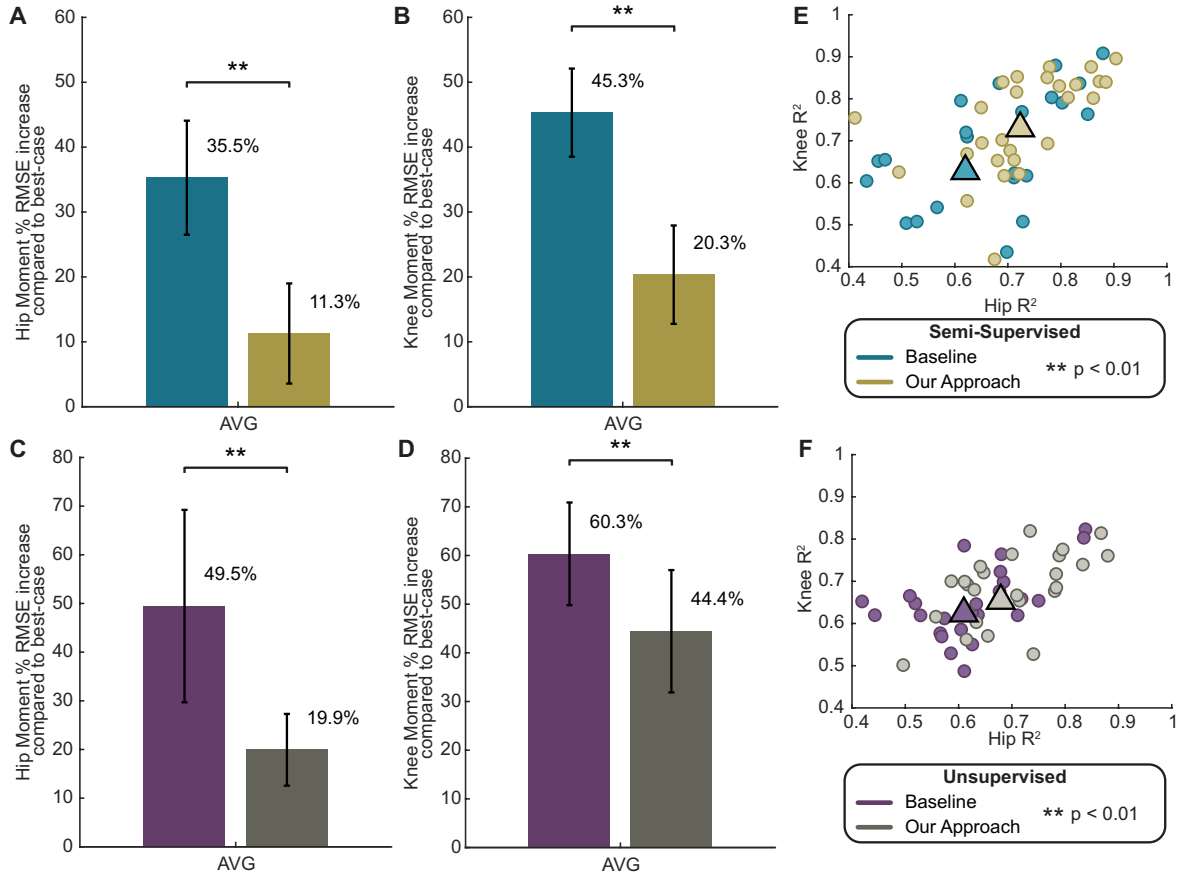

Fig. 3. Offline model performance for moment estimators trained with translated data. For the semisupervised case, the performance of our approach versus the baseline is presented for the hip (**A**) and knee (**B**) in terms of percentage increase in RMSE compared with the best-case model. Parallel results are presented for the unsupervised case at the hip (**C**) and knee (**D**). Averages were first taken across all 28 activities and then across participants. Error bars represent the SD across eight participants, and asterisks represent statistical significance determined by paired *t* tests. The R^2 values between ground truth joint moment and estimate for each of the 28 tasks at the hip and knee are presented for the semisupervised case (**E**) and unsupervised case (**F**). Each point represents the average across eight participants, but error bars were omitted for visual clarity; each triangle represents the average across all tasks.

unsupervised case, our approach significantly outperformed the baseline approach at the hip (percentage increase in RMSE compared with the best-case model of 25.6 ± 13.9% versus the baseline of 63.4 ± 24.8%, $P < 0.01$) and at the knee (percentage increase in RMSE compared with the best-case model of 32.9 ± 6.7% versus the baseline of 54.6 ± 8.1%, $P < 0.01$) (Fig. 4, E and F). The unsupervised baseline approach was computed post hoc because of the poor estimation of the baseline controller causing instability (fig. S3 and movie S1) and deemed unfeasible for real-time performance. The final performance of our unsupervised approach was 0.21 ± 0.01 Nm/kg RMSE and 0.68 ± 0.04 R^2 at the hip and 0.18 ± 0.01 Nm/kg RMSE and 0.71 ± 0.03 R^2 at the knee. The data available to each model and the strategy for using it are depicted in Fig. 4 (G to I).

To understand the effect of these errors on human performance, we deployed the best-case model, our semisupervised model, and our unsupervised model ($N = 8$) during a weightlifting activity and incline walking and compared metabolic cost relative to no exoskeleton (Fig. 5, A and B). There were significant differences between controller types for lifting weight [$F = 8.54$, degrees of freedom (df) = 3, $P < 0.01$] and 5° incline walking ($F = 11.05$, df = 3, $P < 0.01$). Our semisupervised model resulted in statistically significant reductions in metabolic cost compared with no exoskeleton for both lifting a weight and 5° inclines (12.5% for lifting and 14.6% for inclines, both $P < 0.01$). These reductions were similar to the best-case model that also had statistically significant reductions for both tasks (14.4% for lifting and 13.6% for inclines, both $P < 0.01$). In addition, our unsupervised model also reduced metabolic cost compared with no exoskeleton, although it was not statistically significant for eight participants (9.5% for lifting, $P = 0.066$, and 13.8% for inclines, $P = 0.051$). Similar results can be seen between the best-case and our domain adaptation models for tasks that did not require high net-positive work (fig. S8).

To further understand the accuracy of the three models that we deployed in real time, we broke down the results into each specific task. These results are presented in terms of an R^2 shape match between the ground truth joint moment and the estimate (Fig. 6A). The tasks are separated into cyclic, impedance-like, and unstructured activities to demonstrate the performances across all categories (*12*). A one-way analysis of variance (ANOVA) test found statistical differences between the three real-time models for R^2 at the hip ($F = 27.63$, df = 2, $P < 0.01$) and at the knee ($F = 14.47$, df = 2, $P < 0.01$). Follow-up pairwise multiple comparison tests with Bonferroni correction

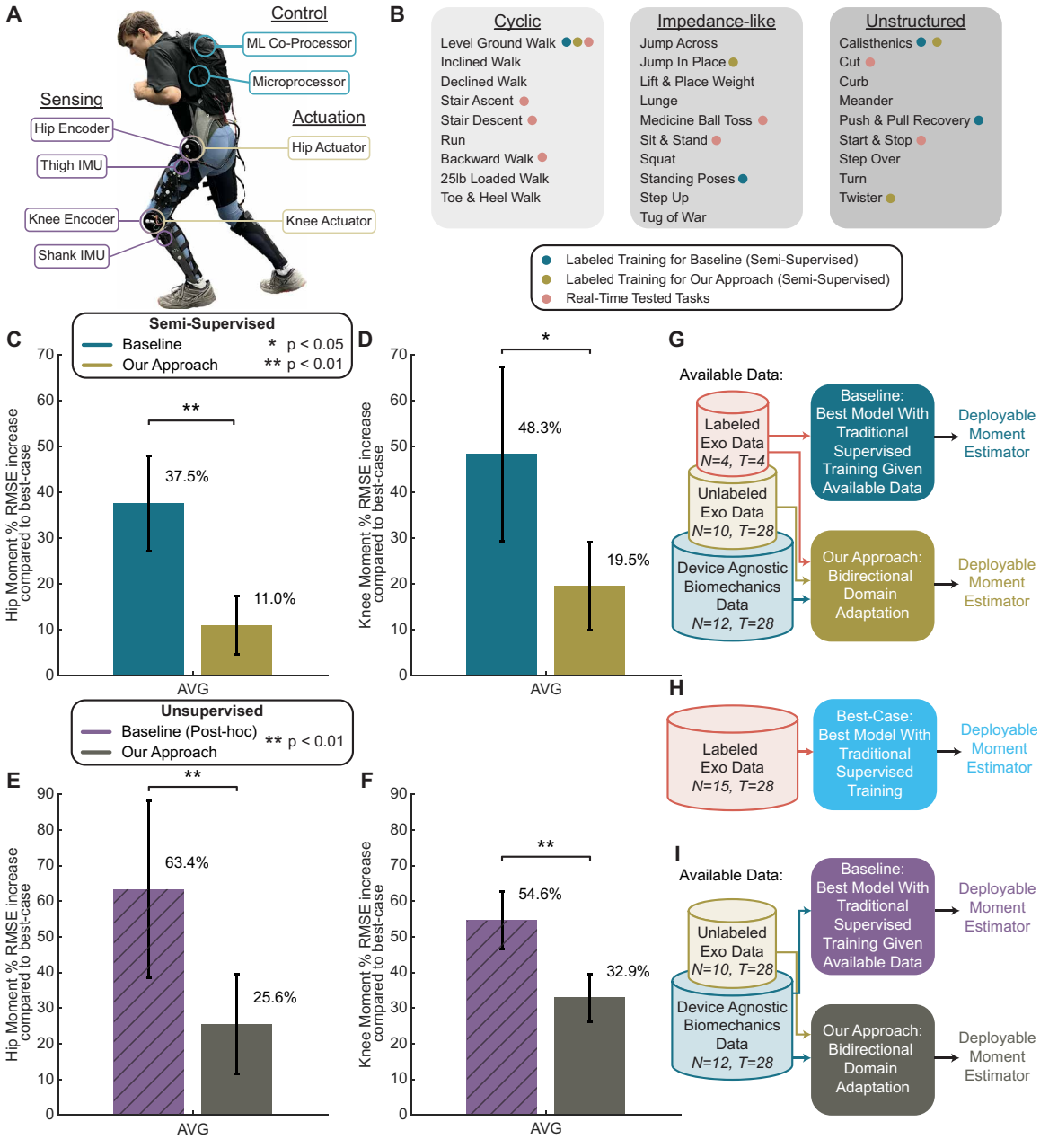

Fig. 4. Real-time model performance for moment estimators trained on translated data. Our autonomous hip/knee exoskeleton (**A**) sends measurements from hip and knee encoders and thigh and shank IMUs through a microprocessor to an onboard machine learning (ML) coprocessor for real-time inference. Inference results converted to actuator commands provide bilateral hip and knee assistance. (**B**) Models with labeled exoskeleton data (semisupervised) included four tasks. Seven real-time tasks were novel to the unsupervised and semisupervised models; level ground walking was in the semisupervised training sets. Unlabeled exoskeleton data and biomechanics datasets included all tasks. The performances of our deployed models versus their baselines are presented for the semisupervised case at the hip (**C**) and knee (**D**) and the unsupervised case at the hip (**E**) and knee (**F**) as a percentage RMSE increase compared with a best-case model (computed post hoc). The unsupervised baseline model was run offline post hoc because of real-time controller instability. Averages were taken across all eight tested tasks and then across eight participants, with error bars representing across-participant SD. Asterisks represent statistical significance determined by paired *t* tests. Pictorial representations of the data for the semisupervised models (**G**), best-case model (**H**), and unsupervised models (**I**) are also presented.

(Fig. 6C) showed a statistically significant increase in R^2 for our semisupervised approach over the semisupervised baseline ($P < 0.01$), our unsupervised approach over the semisupervised baseline ($P = 0.013$), and our semisupervised approach over our unsupervised approach ($P = 0.035$). At the knee, statistically significant improvements in R^2 were found for our semisupervised approach over the semisupervised baseline ($P = 0.010$) and our semisupervised approach over our unsupervised approach ($P = 0.011$).

Figure 7 demonstrates the accurate tracking of joint moments for the semisupervised and unsupervised models through representative

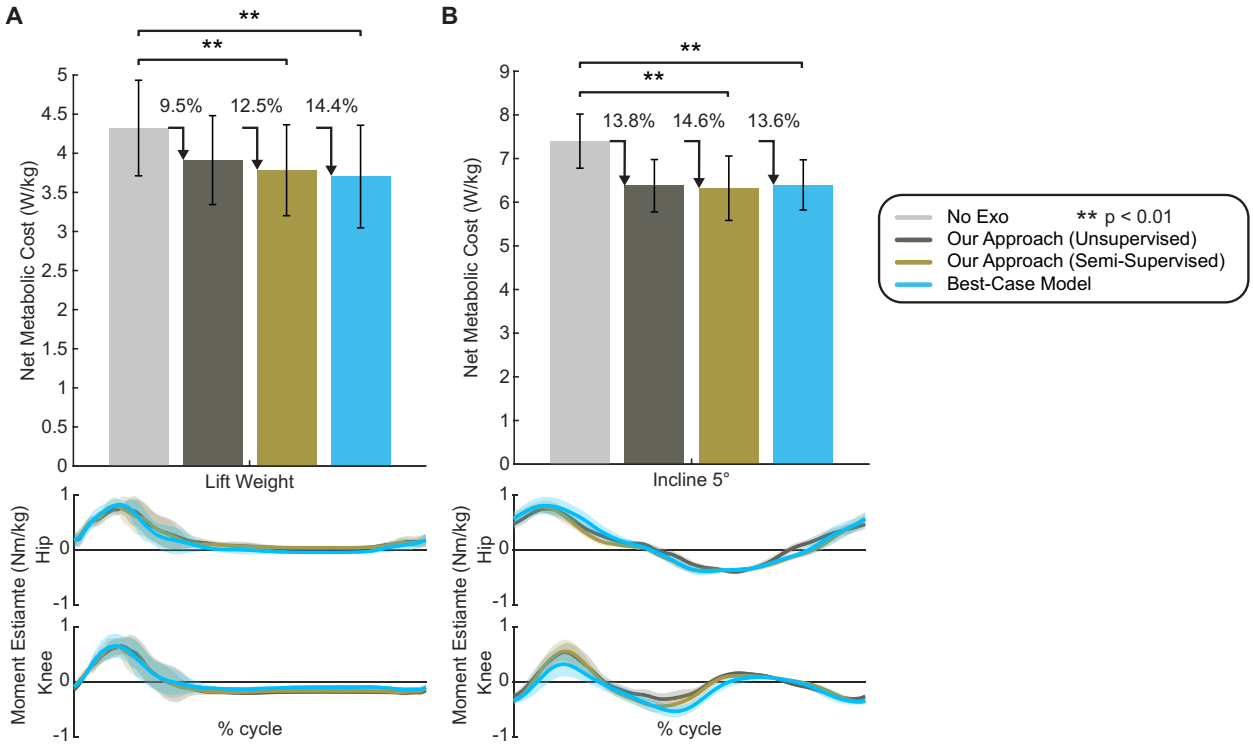

Fig. 5. Net metabolic cost comparisons for moment estimators trained on translated data. Eight participants performed a lift weight task (**A**) and an incline walking task (**B**) with our two controllers and the best-case model as well as without an exoskeleton. Bars represent the average net metabolic cost per body mass, and error bars represent SD across the eight participants. Asterisks indicate statistical significance from follow-up multiple comparison tests with Bonferroni correction ($P < 0.05$). Below the metabolic cost results are cycle-averaged estimates of hip and knee moment for each of the deployed controllers. Real-time estimates are split by cycle and then averaged separately by participant. The haze represents the SD across the eight participants, and the solid line represents the average.

time-series plots from a single cycle of each type of activity. Given that each activity is performed individually with each model, the ground truth moments are not the same for each repetition of the maneuver. The specific strides were selected so that the hip and knee moment RMSE and R^2 are close to the task average and the ground truth moments are also somewhat similar between controllers.

DISCUSSION

These results demonstrate that by leveraging our domain adaptation strategy to exploit data from the biomechanics domain without an exoskeleton, we can create useful data-driven controllers with only limited or even no access to device-specific, labeled data. Although these results are framed in the context of joint moment estimation for task-agnostic control, we showed that the sensor translation can take data without an exoskeleton and replicate characteristics of exoskeleton data, improving downstream models in the real exoskeleton data domain. This is further supported by our results using sensor translation to improve GRF estimation and activity classification. This unified framework for incorporating data from many different sources for real-time deployment has the potential to open previously unrecognized possibilities in device control by creating a large dataset for improving current algorithmic techniques and potentially enabling new, more complex deep learning algorithms.

The strong performance of our approach across a variety of tasks demonstrates its flexibility to improve performance on different size datasets. The task optimization sweep provides a potential look-up table for the number of tasks needed to reach a predefined performance accuracy. There is a tradeoff between the number of labeled tasks and the corresponding estimator performance with diminishing returns as the number of included tasks increases. However, our approach reduces the number of training tasks necessary, so much so that even using nine tasks with the baseline approach has higher error (hip: 26.2%, knee: 29.2%) than our approach with a single task (hip: 18.3%, knee: 19.4%). Thus, our approach makes it feasible to collect far fewer tasks for training. Although the exact optimized tasks differ from those of the baseline, similar trends appear (level ground walk and calisthenics are included in both) emphasizing a diverse set of tasks that covers the extremes of human movement. Similarly, adding labeled participants improves model performance but with decreasing benefit from each additional participant. Including four participants of labeled data achieved moment estimation error within 5% of including all 13 participants; therefore, it is feasible to collect far fewer participants and still achieve comparable results with our approach.

Although these participant and task sweeps involved testing on a left-out participant who was unseen in the training process, aggregate decisions such as hyperparameters were made on the basis of performance on those participants. Thus, the offline performance of our models on the truly novel participant set further confirms that our approach outperforms the baseline approach. When compared with the best-case model with access to all of the labeled data, our semisupervised approach, operating with only 5.1% of the labeled data (525 thousand labels versus 10.3 million labels), only incurred

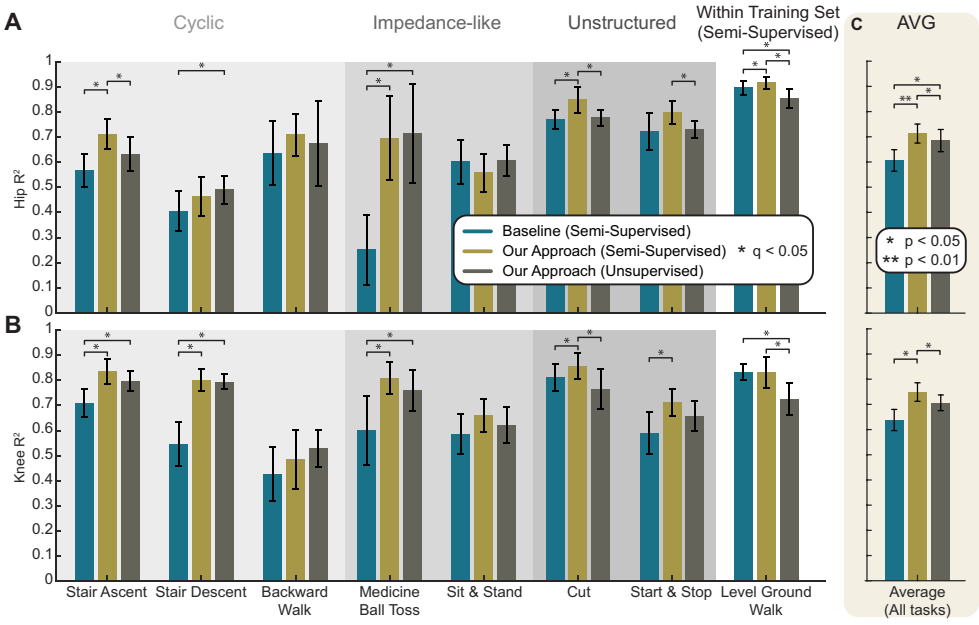

Fig. 6. Real-time deployed model performance broken into specific tasks. The individual R^2 performance of each deployed model on each task is presented for the hip (**A**) and the knee (**B**). Tasks are segmented into four categories. Three are based on the category of activity: cyclic, impedance-like, and unstructured; the fourth is based on the fact that level ground walking was one of the tasks for which the semisupervised models had access to ground truth labels during training, whereas the unsupervised model did not. Statistical significance was assessed using the false rate of discovery ($q < 0.05$) to account for all 24 comparisons. The average R^2 across all tasks (**C**) allows a direct comparison between the three deployed models. Asterisks indicate statistical significance from follow-up multiple comparison tests with Bonferroni correction ($P < 0.05$). Error bars represent the SD across eight participants.

a 10 to 20% increase in moment estimation RMSE at the hip and knee. Our unsupervised model demonstrates the superior performance of our training approach compared with the corresponding baseline using only simulated sensors. This demonstrates the benefits of using unlabeled exoskeleton data to facilitate the translation from simulated to real sensors. Although using no labeled data incurred additional error compared with the best-case model, the performance still remained between 20 and 45% of the best-case moment estimation RMSE at the hip and knee, which led to a stable system in all test scenarios. Although it is difficult to determine which aspects of the unlabeled data the model is learning, our preliminary exploration demonstrated that part of the benefit comes directly from learning to handle real sensor noise and part of the benefit seems to come from understanding the human interacting with the exoskeleton even with the exoskeleton unactuated (fig. S9). Also, by looking at the error reduction as each unlabeled participant was added, we did not see the improvement level off at 10 participants, indicating that more unlabeled data beyond that available in this study could potentially improve these models even further (fig. S10). Overall, these results demonstrate the potential for control engineers without access to a gait laboratory to use control architectures that would otherwise have been impossible.

Our real-time deployment further confirms the offline results and demonstrates that moment estimators trained with our framework maintain offline accuracies while estimating moments in real time (fig. S1). When tested in real time, our semisupervised approach again significantly outperformed the baseline model and incurred only a small penalty relative to the best-case model (between 10 and 20% increase in RMSE across the hip and knee). The same can be seen for the unsupervised model (between 25 and 35% across the hip and the knee). A key factor to note is our inability to run the unsupervised baseline during real time (fig. S3 and movie S1). This demonstrates that our approach takes a previously impossible controller, trained only on simulated data, and makes it possible and useful in the real world. Not only was this viable, but across tasks, our unsupervised approach even outperformed the semisupervised baseline (statistically significant at the hip, but not at the knee). The only case where it performed statistically worse than the semisupervised baseline was during level ground walking, which is reasonable given that this was the only tested task that was part of the labeled training set for the semisupervised baseline. The task breakdown (Fig. 6) demonstrates the additional benefit of including labeled data within our framework, as evidenced by statistically significant differences between our semisupervised approach and our unsupervised approach.

Although previous studies have demonstrated beneficial human outcomes from using moment estimation–based controllers (*11*, *12*), the metabolic cost results presented here demonstrate that models trained with translated data can also provide energetically beneficial assistance to users even on high torque tasks. Our semisupervised models reduced metabolic cost by a similar percentage as the best-case model, indicating that the small increase in estimation error had very little effect on performance for these tasks. Although there is some loss in consistency for metabolic performance without any labeled data, the model still enabled sizable reductions in metabolic cost for users across these tasks. This is unexpected because without our translation approach, the baseline comparison could not even be used in real time. These results demonstrate that our models overcame the metabolic cost penalty of wearing our 7-kg device and provided meaningful assistance to augment users' performances. Reductions greater than 10% are comparable to the current state-of-the-art controllers (*32*).

The unsupervised result demonstrates that it is possible to take labeled data from readily available biomechanics datasets combined with unlabeled data from a novel device and create a fully deployable deep learning network for that new device. Thus, researchers with minimal access to a gait lab can use open-source data, possibly combined from many different sources, to form a large set of labeled simulated exoskeleton data. Then, using that data with some additional data from users moving in a controllerless novel device outside the research laboratory, they can train a fully deployable task-agnostic controller. Given that the framework presented here performs sensor translation first and then downstream moment estimation, this same framework can be applied to many different types of data-driven

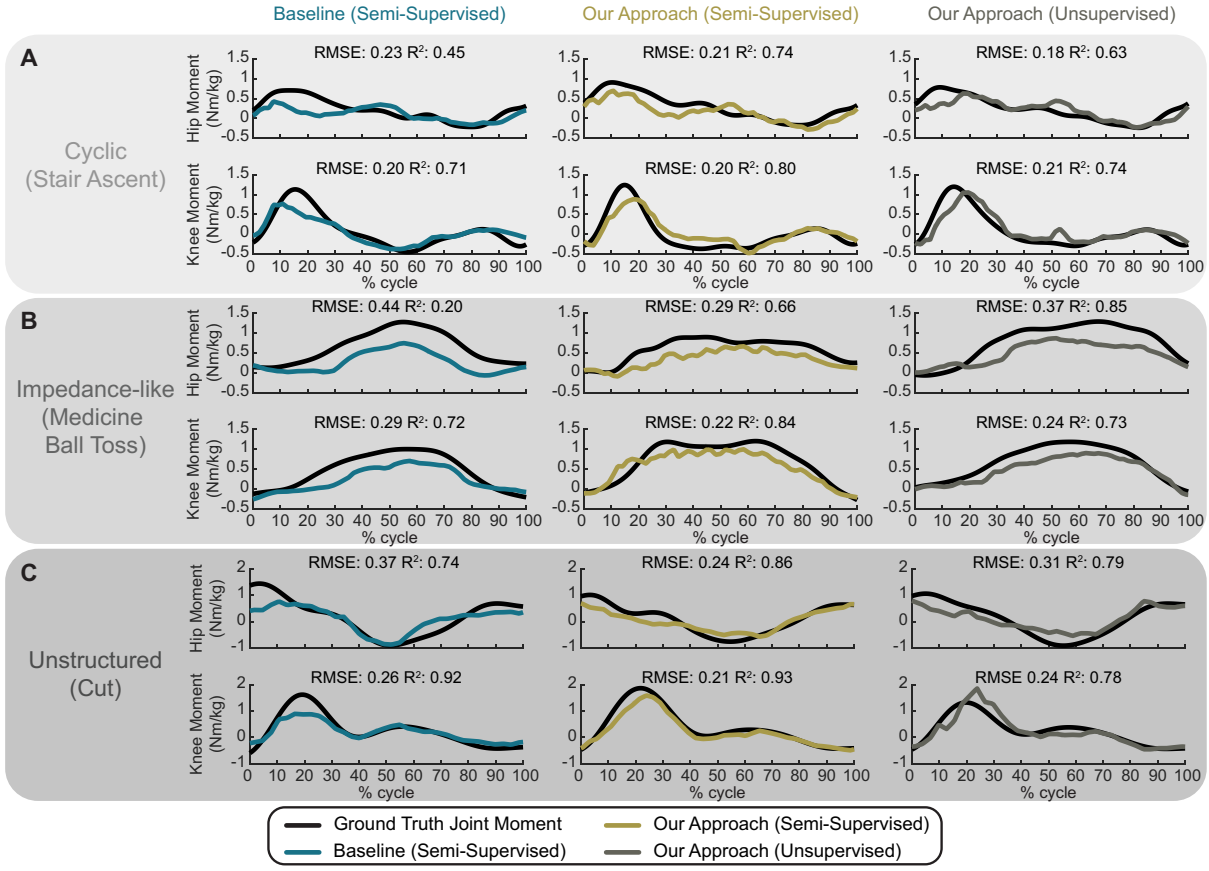

Fig. 7. Time series comparison of deployed models. Representative time-series plots are shown for each category of activity with the ground truth moment in black and the real-time joint moment estimate in color. Stair ascent is representative of cyclic tasks (**A**), medicine ball toss of impedance-like activities (**B**), and the outside leg for a single stride in cutting for unstructured tasks (**C**). To facilitate comparison across controllers, individual cycles were chosen by finding the three strides where the hip and knee RMSE and R^2 most closely matched the task average and then selecting the final cycles on the basis of ground truth moment similarity.

models beyond joint moment estimation, as we showed with GRF estimation and activity classification. This has the potential to enable faster development and broader use of these powerful techniques.

Our semisupervised approach provides a roadmap for continuing to improve these data-driven approaches over time. As more data become available through testing with the initial deployed model, these data can be efficiently exploited to continue to improve model performance. For researchers who already have some labeled data with their device, these results demonstrate a method for incorporating that data into the model. This work shows that with just four tasks from four participants—a tiny fraction of our overall dataset—estimation error with respect to the best-case model can be cut by half compared with having no labeled data.

We propose that our domain adaptation approach is applicable across many areas of exoskeleton research. Our task optimization results and those from Molinaro *et al.* (*12*) demonstrate that having data from a specific activity continues improving data-driven approaches for that specific activity even if the additional data do not promote generalization to other activities. Thus, if researchers and designers want to add additional functionality for a specific task (e.g., a golf swing), at present, they would have to collect exoskeleton data with that specific task. However, with the framework presented here, these data could come from an open-source biomechanics dataset without incurring data collection and labeling costs. Furthermore, this approach could be a valuable tool for creating models for unique populations. It is difficult to collect large participant datasets of movement for elderly individuals or for those with impaired mobility while wearing a device, but some datasets without a device are readily available (*33*–*35*). Hence, our framework could potentially make use of those datasets for use on new devices targeted toward these populations.

Limitations

This study has several limitations. First, our analysis of the unlabeled data demonstrated that we had not saturated performance with 10 participants. More benefit may be found from including additional participants' unlabeled data and selectively adding or removing unlabeled tasks. Second, only IMUs and encoders were used in translation to avoid using sensors outside the device. Exploring other sensing modalities beyond kinematics was beyond the scope of this study. Third, we performed initial optimization of individual network hyperparameters on a network-by-network basis; however, additional gain may be found by optimizing those hyperparameters in the context of the entire network. Fourth, although there is nothing device specific in our translation method, we only tested our models on the hip/knee exoskeleton presented. Last, we did not examine the

ability of our domain adaptation models to generalize to completely novel tasks that are not in the unlabeled dataset or open-source biomechanics datasets because of the relative ease of collecting or obtaining these data.

Conclusion

This paper presents our framework that considers the simulated sensor domain (on the basis of biomechanical models) as a stepping-stone domain for aggregating data. This allowed us to train and optimize a deep domain adaptation network that translates data from large biomechanics datasets into a device-specific, real sensor domain where labels are difficult and costly to collect. These translated data can be used for training downstream deep learning models, such as joint moment estimators, that allow powerful task-agnostic and user-independent exoskeleton control strategies. Using this network, we trained models using both no labeled data (unsupervised) and a small-scale dataset (semisupervised) for a real hip/knee exoskeleton. These models incurred only small additional estimation error compared with a best-case model with a complete labeled dataset from our exoskeleton and significantly outperformed models trained without domain adaptation. We also proved that these models could be deployed in real-time exoskeleton control, achieve comparable performance to the offline models, and augment human performance by reducing metabolic cost. This represents a breakthrough in user accessibility to data-driven approaches for exoskeleton control by cutting the need for costly data collection, a key weakness of our former approach (*11*, *12*), and providing a method to incorporate readily available human biomechanics data to improve model performance.

MATERIALS AND METHODS

Framework overview

The goal of our approach was to decrease the dependence on costly labeled data by replacing those data with easily accessible data. To make these accessible data useful, we translated data (sensor signals) that had associated labels (in our case joint moments) from a domain where the labels were easy to obtain through a stepping-stone domain (simulated sensors) into the data-limited domain (the real exoskeleton domain). We accomplished this using a translator that was trained through a CycleGAN with U-Net backbones. In this work, we considered the following three sources of data, each with a different level of costliness to collect and each allowing access to simulated sensors, real sensors, or both (Fig. 1). Labeled human biomechanics data collected without an exoskeleton are the least costly to obtain (our source domain) and can easily be transformed into our stepping-stone domain by simulating sensors similar to the device (dataset $\mathbb{D}_{ls}$); however, these data are the least valuable for direct use because they do not capture the richness of real sensor inputs. Unlabeled, unpowered exoskeleton data (unlabeled data in our target domain) are collected from a human user wearing a specific target device but without any need for external measuring equipment or a stand-in controller providing assistance (dataset $\mathbb{D}_{ur}$). Thus, these data are relatively easy to obtain in any environment but lie solely in the real domain. The third source, labeled exoskeleton data, is the costliest data to collect because they require a gait laboratory with instrumented floors and motion capture; however, they contain time-synced information from both the real sensor domain and our stepping-stone simulated sensor domain and therefore are easy to use to create a translation between the two domains (dataset $\mathbb{D}_{lr}$). Our framework uses previously underused data from the first two less costly sources to reduce or even remove the need for the third costly data source, which is traditionally the only source used for deployed machine learning algorithms (*3*, *11*).

To achieve real-time deployable models for exoskeleton control, we propose a two-part approach. First, we tackled the domain adaptation problem by training two translator networks through respective GANs, one to convert from the simulated to the real domain ($T_{s\to r}$) and vice versa ($T_{r\to s}$). For translation, we had access to many unpaired samples from the simulated sensor domain (X_s) (drawn from our source human biomechanics dataset domain) and the real sensor domain (X_r) (unlabeled data from the target domain). Although there is no direct time point–by–time point comparison for conventional supervised training, these unpaired samples can be used through domain adaptation to indirectly learn the translation. Our approach, inspired by a CycleGAN, uses adversarial learning to learn to match the data distribution of the translated sensor signals with that of the real sensor signals in the target domain. The unpaired samples from the simulated and real domains are the only data accessible in the unsupervised case.

In the semisupervised case, we had limited access to time-paired data samples from both the real (X_r) and simulated domain (X_s) (labeled exoskeleton data). However, we could not just learn a translation on the basis of these data because the semisupervised time-paired data are only available while wearing the exoskeleton. Thus, although any translation trained on these time-paired data would capture the noise characteristics of real sensors and any orientation differences between the simulated and real sensors, it would fail to capture the user's kinematic changes because of wearing the device (whether from human adaptation to assistance or the exoskeleton mass and range of motion constraints). Those changes are something we wanted the translator to learn. Thus, these data provided a few samples with labels that partially constrain, although imperfectly, the translation ($T_{r\to s}$ and $T_{s\to r}$) but could not perform as well on their own (fig. S2) or allow a direct measure of translator performance.

In the second part of our approach, we used one of the translator networks from domain adaptation ($T_{s\to r}$) to translate data from the generic human biomechanics datasets (the common shared domain X_s) into the real domain. Because these data came from datasets that contain inverse dynamics, the appropriate joint moment labels associated with the data were available (y_s). These label-paired data samples {i.e., translated sensor data $[T_{s\to r}(X_s)]$ and joint moment labels (y_s)} could then be used to train a moment estimator that could operate based on data from the real device. If we had access to some real device data with joint moment labels (semisupervised case X_r, y_r), then these could also be added into the training set for the moment estimator.

To summarize, we used three datasets: Dataset $\mathbb{D}_{ls}$ contained labeled human biomechanics data without an exoskeleton containing simulated sensor signals (X_s) and time-matched joint moment labels (y_s). Dataset $\mathbb{D}_{ur}$ contained unlabeled real sensor data with an exoskeleton containing real sensor signals (X_r) and no labels. Dataset $\mathbb{D}_{lr}$ contained labeled real and simulated sensor data with an exoskeleton containing real sensor signals (X_r), time-matched simulated sensor signals (X_s), and time-matched joint moment labels (y_r).

Domain adaptation network details

The sensor translation portion of the network combined two GANs, one generating data in the real sensor domain conditioned on data

from the simulated domain ($T_{r\to s}$) and the other operating in reverse ($T_{s\to r}$). Each GAN consisted of a U-Net–style translator (T) as the generator and a convolutional neural network (CNN) as the discriminator (D).

The GANs were set up bidirectionally, following a CycleGAN framework (*36*). Results from image translation and human activity recognition have demonstrated that bidirectional translation increases performance by encouraging minimal information loss across the cycle and forcing the translations to be inverses of each other (*29*, *36*, *37*). For moment estimation *ME*, a temporal convolutional network (TCN) architecture was selected on the basis of its success in previous studies (*11*, *12*, *38*).

Given the bidirectional nature of the translation, the moment estimator could be trained to operate based on real data (ME_r) or on simulated data (ME_s). We found that training the moment estimator on the real side (ME_r) outperformed the moment estimator on the simulated side (ME_s) with the added benefit of only requiring one model to be deployed in real time (for real-time implementation, the moment estimator trained on the simulated side would require both the real-to-simulation translator and the moment estimator to operate on incoming data). Thus, all results are based on a moment estimator trained with data that are either converted to or already in the real domain (ME_r). The networks are depicted in Fig. 8 along with the flow of data necessary to compute and combine multiple training losses. Five different types of losses were available to differing degrees based on the specific case, semisupervised or unsupervised. The rationale and equations for each loss function, the optimization and tuning of the hyperparameters for the overall framework, the architecture selection, and further hyperparameter tuning for the U-Net, TCN, and CNN architectures are described in Supplementary Methods and tables S1 and S2. Further details about training procedures are also included there. An in-depth analysis of the value of the GAN loss is provided in the Supplementary Materials and fig. S2.

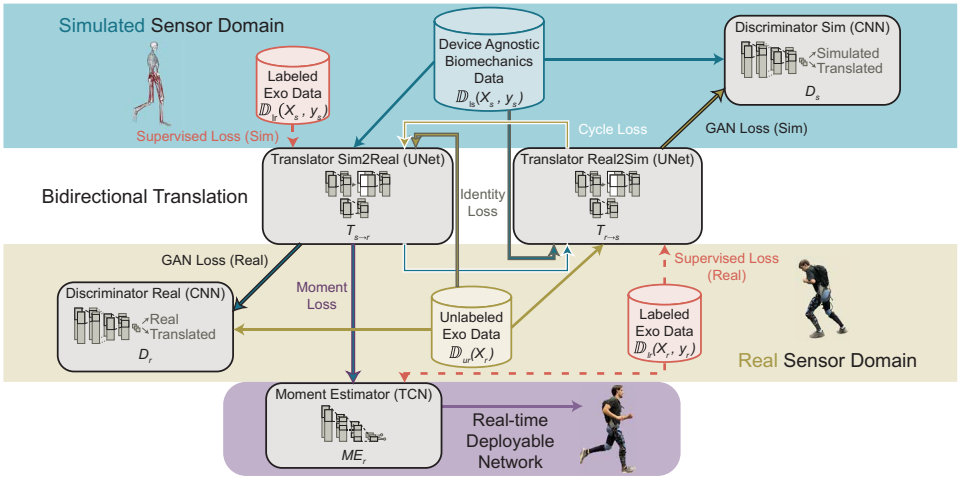

Fig. 8. Detailed network description for training the translator and downstream moment estimator. Bidirectional domain adaptation is accomplished through two pairs of translator and discriminator networks that function as GANs in both simulated and real domains. The forward pass of data is pictured by the arrows with the color indicating the original data source. The outline colors depict the portions of the forward pass that contribute to different components of the loss function including supervised loss, cycle loss, GAN loss, identity loss, and moment loss. The final fifth network, the moment estimator, is trained along with the other components but is the only portion required at run time for deployment on the device.

Training datasets

In training our framework, we chose to estimate hip and knee moments in keeping with Molinaro *et al.* (*12*). For exoskeleton sensors, we chose to use purely kinematic sensors on the thigh and shank. Thus, our sensor suite consisted of a hip and knee encoder (angle and velocity) and a thigh and shank inertial measurement unit (IMU). This mimics the most basic set of mechanical sensors available to a hip/knee exoskeleton without placing sensors outside the device.

Labeled biomechanics data (without an exoskeleton $\mathbb{D}_{ls}$) came from Scherpereel *et al.* (*39*). These data included joint angles, joint moments, and simulated IMUs on each segment for 12 participants performing 28 cyclic and noncyclic groups of tasks. We simulated encoders from the hip and knee inverse kinematics and simulated IMUs for the thigh and shank, matching IMU locations with those on our hip/knee exoskeleton.

Exoskeleton data came from Molinaro *et al.* (*12*), which included several phases of collection with 22 different participants. Our unlabeled real exoskeleton data ($\mathbb{D}_{ur}$) consisted of raw sensor data collected from the exoskeleton, whereas 10 participants performed tasks without exoskeleton assistance (actuators off). We used the real IMU data for the thigh and shank (OpenIMU, Tewksbury, MA) and encoder data from the hip and knee (T-Motor AK80-9s, Nanchang, China). For our semisupervised model, we included a combination of actuated and unactuated data from the first 15 participants across a range of tasks ($\mathbb{D}_{lr}$). Eight participants from "phase 3" of this dataset formed our true test set and were used to evaluate performance. No data from these participants were used to train the models ensuring that these were novel participants to verify that the models were user independent. Additional details on modifications to this dataset are described in the Supplementary Materials.

Assessing the joint moment estimation performance of the model

We assessed the estimator performance using RMSE between the model estimated moments and the actual ground truth joint moments calculated using inverse dynamics from OpenSim. We then averaged the RMSE and R^2 across subtasks in the larger 28 task groups and then across tested participants (*12*).

To provide suitable comparisons to demonstrate the performance of our algorithm, we present the results relative to two models: a best-case model (our goal being to achieve accuracies similar to those of this model) and a baseline model, which was trained without domain adaptation. To construct the best-case comparison, we trained a moment estimator using all of the training tasks and participants, thus replicating the results of Molinaro *et al.* (*12*) but without pressure insoles, a foot IMU, or delayed hip moment estimates and using all 28 tasks for training (Fig. 4H). Although it would be theoretically possible given a substantially larger set of no exoskeleton data to achieve higher accuracies with our method than this best-case model,

our no exoskeleton dataset was of similar size. For the baseline model in the semisupervised case, we trained a moment estimator in the same way as the best-case model but with only the same limited number of tasks and limited number of participants that we gave to our own approach (Fig. 4G). The baseline model in the unsupervised case was trained with only simulated sensor data for participants without an exoskeleton (10 participants and 28 tasks as in Fig. 4I). These baselines were the best models that could be constructed using traditional supervised learning techniques given the data limitations.

To report our results, we computed the percentage increase in error for our algorithm and for the baseline with respect to the best-case model as seen in Eq. 1

$$\%\,\text{increase in RMSE} = \frac{\text{RMSE}_{\text{approach}} - \text{RMSE}_{\text{bestcase}}}{\text{RMSE}_{\text{bestcase}}} \cdot 100\,\% \quad (1)$$

The value of this percentage increase expresses how far the approach is away from the best case, and the percentages can be compared between our approach and the baseline.

Real-time experimental validation

In contrast with offline performance (where previously collected and held-out data are passed through the model), validating real-time performance required us to deploy three models in our real-time controller loop. For the semisupervised case, we chose to deploy both the baseline and our approach in real time. As explained in the Results section, four tasks and four participants were selected for use in the labeled training set. The specific four tasks were based on their respective optimal task ranking. For the unsupervised case, we deployed only a model trained with our approach. Although we attempted to deploy the baseline model for this case, the model accuracy with only simulated data was so poor that the controller was not stable for real-time deployment (see movie S1 and fig. S3). To provide a direct offline comparison, we tested the unsupervised baseline model post hoc with the same data collected with our real-time unsupervised model.

To deploy the models in real time, we used the autonomous hip/knee exoskeleton developed by X, the Moonshot Factory, and detailed in our previous publication (*12*). Because the moment estimators are directly real-time deployable, no additional computational time loss was incurred by running our models compared to our previous work. As in our previous work, a main single board computer (Raspberry Pi, Cambridge, UK) handled the input/output for the device and data logging, whereas a separate coprocessor handled deep learning inference (NVIDIA Jetson Nano, Santa Clara, CA). All experimenter interaction was handled through an offboard laptop connected to a Wi-Fi network hosted by the onboard microprocessor. The device has four actuators (AK809 T-Motor, Nanchang), two at the hips and two at the knees for bilateral sagittal plane assistance at hip and knee. The device and mechatronics are pictured in Fig. 4A. As in our previous work, the estimated biological moment was converted to applied torque by first scaling the moment (by participant mass and then by a scaling factor of 20% for the hip and 15% for the knee), then delaying the torque (by 100 ms for the hip and 50 ms for the knee), and last filtering the torque using a low-pass filter (second-order Butterworth with a 10-Hz cutoff frequency) (*12*).

We performed a protocol similar to that used to collect the training data (*12*, *39*) with a pared-down task list. Eight novel participants not included in the training or offline testing data (four males, four females; age: 23.0 ± 3.8 years; height: 173.4 ± 2.6 cm; body mass: 68.2 ± 10.7 kg) performed eight tasks with three different controllers. Informed consent was obtained from each participant under Georgia Institute of Technology Institutional Review Board protocol H21184. The tasks were split between the three larger task categories used in (*12*) with seven tasks outside of the labeled training set for the semisupervised models as seen in Fig. 4B. These tasks were medicine ball toss [15 lb (6.8 kg) center, right, left], cutting (left and right, fast and slow), sit-to-stand (short and tall chair, with and without armrests), starting and stopping, stair ascent (slow, normal, fast), stair descent (slow, normal, fast), and walking backward (0.6, 0.8, 1.0 m/s). More details on the protocols for each task can be found in our previous dataset publication (*39*). We also included one task that was in the training set for the semisupervised models: level ground walking (0.6 m/s, 1.2 m/s, 1.8 m/s, and shuffle). The order of the models was randomized and blinded for each participant; all tasks were performed with a single model before switching models. During each task, exoskeleton data for the encoders (hip and knee), IMUs (thigh and shank), and the estimated joint moments from the current model were recorded. Motion capture data were recorded at 200 Hz and GFRs at 1000 Hz.

Data processing was performed in a similar manner to our previous work (*12*) where participant-specific musculoskeletal models were fit to each participant standing in a static pose. The exoskeleton mass was then added for the actuators at their respective joints and the additional mass was added at the torso. Inverse kinematics matched the model to the measured motion capture locations, and inverse dynamics were then calculated on the basis of the GFRs using OpenSim (*30*, *31*). An activity flag was computed to determine when the activity was occurring and then used as the sections to assess model performance. Ground truth joint moments were filtered before being compared with the real-time estimates. Further details on biomechanical processing and analysis can be found in our previous work (*12*, *39*).

To assess the real-time model performance, we captured the joint moment estimates and then post hoc aligned those estimates with the ground truth moment labels as in (*12*). We did not deploy the best-case model in real time when collecting data for joint moment accuracy but instead ran the collected sensor signals through the best-case model post hoc, thus allowing a fairer comparison of performance. We did this individually for the data from each model; thus, the best-case model performance was unique to each specific model's data (except in the case of the unsupervised baseline as noted above).

We further assessed human performance outcomes by having eight novel participants collect metabolic cost measurements during a lifting task (six males, two females; age: 25.0 ± 3.9 years; height: 174.6 ± 2.4 cm; body mass: 71.0 ± 9.8 kg) and an incline walking task (five males, three females; age: 25.0 ± 2.1 years; height: 175.3 ± 4.2 cm; body mass: 68.9 ± 7.0 kg). Before any metabolic cost collection, we performed the same acclimation procedure as (*12*). For the lifting task, a metronome was set at 10 beats per minute, and participants were instructed on each tone to use both hands to lift a 25 lb. (11.3 kg) kettle bell weight off a waist-height shelf, touch that weight to the floor between their feet, and then replace the weight on the shelf (*12*). For the incline task, participants walked on a 5° inclined treadmill at 1.25 m/s. Four conditions were tested for each activity: our semisupervised model, our unsupervised model, the best-case model, and no exoskeleton. Participants performed the activity

for 6 min, and conditions were tested using a within-participant counter-balanced design (ABCD-DCBA). Because of an exoskeleton malfunction, one incline trial for one participant only lasted 3 min, so a first-order fit estimate was used for that trial (*40*, *41*) and averaged with the other 6-min trial with the same controller. The order of the three exoskeleton conditions was randomized, and participants were blinded to the condition; however, for a given task, the no exoskeleton condition was placed either at the beginning (A) or the end (D) for don and doff efficiency. Motion capture—and thus ground truth joint moments—was not available during metabolic cost collections. Results for all metabolic cost measurements are reported on the basis of an average of instantaneous metabolic cost (as calculated with the Brockway equation) for the final 3 min of each activity.

Statistical analyses

All statistical analyses were performed in MATLAB (MathWorks, Natick, MA) using a significance level of $\alpha = 0.05$. For the task and participant performance sweeps, significance was assessed by controlling the false rate of discovery ($q < 0.05$) across model conditions (our approach versus the baseline), where percentage RMSE increase compared with the best case was the dependent variable (*42*, *43*). We assessed the significance of our approach versus the baseline with each added task or participant and used the false rate of discovery to control the family-wise error rate across all comparisons (9 comparisons for the task case and 13 for the participant case).

For the final offline testing set, we assessed significance using paired *t* tests between the baseline and our approach at the hip and separately at the knee. Percentage RMSE increase compared with the best case was the dependent variable, which was averaged across task groups. The independent variable was the model type; participants were the fixed effect.

For real-time testing, comparisons between the baseline and our approach were again performed with paired *t* tests separately for hip and knee and for the supervised and unsupervised cases. Tests were performed with percentage RMSE increase compared with the best case as the dependent variable. Again, the independent variable was the model type, and participants were the fixed effect. For metabolic comparisons, we performed repeated one-way analysis of variance (ANOVA) tests for each task with participants as the random effect. This was followed up by planned Bonferroni multiple comparison tests to compare each controller to the no exoskeleton condition but not to each other.

To assess the statistical significance for each individual task with R^2 as the dependent variable, we again controlled the false rate of discovery to account for the number of comparisons between model types (three models were run in real time) and number of tasks (eight performed in the real-time validation), thus controlling for all 24 comparisons. Tests were performed with R^2 as the dependent variable separately for each model type with offline versus real time as the independent variable. We also compared the average performance across all tasks directly between the three models that were run in real time using a one-way analysis of variance test with participants as the random effect. This was followed up by Bonferroni multiple comparison tests between model types. To assess offline versus real-time performance, separate two-sample *t* tests with assumed equal variance were performed for each deployed model at each joint to determine whether running the model in real time had a statistically significant effect on performance.

Supplementary Materials

The PDF file includes:
Methods
Figs. S1 to S10
Tables S1 and S2
References (*44*–*53*)

Other Supplementary Material for this manuscript includes the following:
Movie S1
MDAR Reproducibility Checklist

REFERENCES AND NOTES

1. R. M. Karulkar, P. M. Wensing, Personalized estimation of intended gait speed for lower-limb exoskeleton users via data augmentation using mutual information. *IEEE Robot. Autom. Lett.* **7**, 9723–9730 (2022).
2. I. Kang, P. Kunapuli, H. Hsu, A. J. Young, "Electromyography (EMG) signal contributions in speed and slope estimation using robotic exoskeletons," in *2019 IEEE 16th International Conference on Rehabilitation Robotics (ICORR)* (IEEE, 2019), pp. 548–553.
3. M. K. Shepherd, D. D. Molinaro, G. S. Sawicki, A. J. Young, Deep learning enables exoboot control to augment variable-speed walking. *IEEE Robot. Autom. Lett.* **7**, 3571–3577 (2022).
4. I. Kang, P. Kunapuli, A. J. Young, Real-time neural network-based gait phase estimation using a robotic hip exoskeleton. *IEEE Trans. Med. Robot. Bionics* **2**, 28–37 (2020).
5. I. Kang, D. D. Molinaro, S. Duggal, Y. Chen, P. Kunapuli, A. J. Young, Real-time gait phase estimation for robotic hip exoskeleton control during multimodal locomotion. *IEEE Robot. Autom. Lett.* **6**, 3491–3497 (2021).
6. J. Wang, D. Wu, Y. Gao, X. Wang, X. Li, G. Xu, W. Dong, Integral real-time locomotion mode recognition based on GA-CNN for lower limb exoskeleton. *J. Bionic. Eng.* **19**, 1359–1373 (2022).
7. Y. Qian, Y. Wang, C. Chen, J. Xiong, Y. Leng, H. Yu, C. Fu, Predictive locomotion mode recognition and accurate gait phase estimation for hip exoskeleton on various terrains. *IEEE Robot. Autom. Lett.* **7**, 6439–6446 (2022).
8. H. Zhao, Z. Qiu, D. Peng, F. Wang, Z. Wang, S. Qiu, X. Shi, Q. Chu, Prediction of joint angles based on human lower limb surface electromyography. *Sensors (Basel)* **23**, 5404 (2023).
9. T. Lee, I. Kim, S.-H. Lee, Estimation of the continuous walking angle of knee and ankle (talocrural joint, subtalar joint) of a lower-limb exoskeleton robot using a neural network. *Sensors* **21**, 2807 (2021).
10. J. Lin, N. V. Divekar, G. C. Thomas, R. D. Gregg, Optimally biomimetic passivity-based control of a lower-limb exoskeleton over the primary activities of daily life. *IEEE Open J. Control Syst.* **1**, 15–28 (2022).
11. D. D. Molinaro, I. Kang, A. J. Young, Estimating human joint moments unifies exoskeleton control, reducing user effort. *Sci. Robot.* **9**, eadi8852 (2024).
12. D. D. Molinaro, K. L. Scherpereel, E. B. Schonhaut, G. Evangelopoulos, M. K. Shepherd, A. J. Young, Task-agnostic exoskeleton control via biological joint moment estimation. *Nature* **635**, 337–344 (2024).
13. D. Lee, I. Kang, D. D. Molinaro, A. Yu, A. J. Young, Real-time user-independent slope prediction using deep learning for modulation of robotic knee exoskeleton assistance. *IEEE Robot. Autom. Lett.* **6**, 3995–4000 (2021).
14. F. Yang, C. Chen, Z. Wang, H. Chen, Y. Liu, G. Li, X. Wu, ViT-based terrain recognition system for wearable soft exosuit. *Biomim. Intell. Robot.* **3**, 100087 (2023).
15. B. Laschowski, W. McNally, A. Wong, J. McPhee, "Preliminary design of an environment recognition system for controlling robotic lower-limb prostheses and exoskeletons," in *2019 IEEE 16th International Conference on Rehabilitation Robotics (ICORR)* (IEEE, 2019), pp. 868–873.
16. P. Slade, M. J. Kochenderfer, S. L. Delp, S. H. Collins, Personalizing exoskeleton assistance while walking in the real world. *Nature* **610**, 277–282 (2022).
17. J. M. Lopes, J. Figueiredo, P. Fonseca, J. J. Cerqueira, J. P. Vilas-Boas, C. P. Santos, Deep learning-based energy expenditure estimation in assisted and non-assisted gait using inertial, EMG, and heart rate wearable sensors. *Sensors* **22**, 7913 (2022).
18. U. H. Lee, V. S. Shetty, P. W. Franks, J. Tan, G. Evangelopoulos, S. Ha, E. J. Rouse, User preference optimization for control of ankle exoskeletons using sample efficient active learning. *Sci. Robot.* **8**, eadg3705 (2023).
19. L. Zhang, D. Soselia, R. Wang, E. M. Gutierrez-Farewik, Lower-limb joint torque prediction using LSTM neural networks and transfer learning. *IEEE Trans. Neural Syst. Rehabil. Eng.* **30**, 600–609 (2022).
20. B. X. W. Liew, D. Rügamer, X. Zhai, Y. Wang, S. Morris, K. Netto, Comparing shallow, deep, and transfer learning in predicting joint moments in running. *J. Biomech.* **129**, 110820 (2021).
21. J. Sloboda, P. Stegall, R. J. McKindles, L. Stirling, H. C. Siu, "Utility of inter-subject transfer learning for wearable-sensor-based joint torque prediction models," in *2021 43rd Annual*

International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (IEEE, 2021), pp. 4901–4907.
22. Y. Wang, Z. Li, X. Wang, H. Yu, W. Liao, D. Arifoglu, Human gait data augmentation and trajectory prediction for lower-limb rehabilitation robot control using GANs and attention mechanism. *Machines* **9**, 367 (2021).
23. M. Kim, L. J. Hargrove, Generating synthetic gait patterns based on benchmark datasets for controlling prosthetic legs. *J. Neuroeng. Rehabil.* **20**, 115 (2023).
24. Y. Chang, A. Mathur, A. Isopoussu, J. Song, F. Kawsar, A systematic study of unsupervised domain adaptation for robust human-activity recognition. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **4**, 1–30 (2020).
25. F. Mu, X. Gu, Y. Guo, B. Lo, "Unsupervised domain adaptation for position-independent IMU based gait analysis," in *2020 IEEE SENSORS* (IEEE, 2020), pp. 1–4.
26. A. Z. M. Faridee, A. Chakma, A. Misra, N. Roy, STranGAN: Adversarially-learnt spatial transformer for scalable human activity recognition. *Smart Health* **23**, 100226 (2022).
27. C. Hegde, G. Wen, L. C. Price, Activity classification using unsupervised domain transfer from body worn sensors. *Smart Health* **30**, 100431 (2023).
28. A. Akbari, R. Jafari, "Transferring activity recognition models for new wearable sensors with deep generative domain adaptation," in *Proceedings of the 18th International Conference on Information Processing in Sensor Networks* (Association for Computing Machinery, 2019), pp. 85–96.
29. C. Chen, Y. Miao, C. X. Lu, L. Xie, P. Blunsom, A. Markham, N. Trigoni, MotionTransformer: Transferring neural inertial tracking between domains. *Proc. AAAI Conf. Artif. Intell.* **33**, 8009–8016 (2019).
30. S. L. Delp, F. C. Anderson, A. S. Arnold, P. Loan, A. Habib, C. T. John, E. Guendelman, D. G. Thelen, OpenSim: Open-source software to create and analyze dynamic simulations of movement. *IEEE Trans. Biomed. Eng.* **54**, 1940–1950 (2007).
31. A. Seth, J. L. Hicks, T. K. Uchida, A. Habib, C. L. Dembia, J. J. Dunne, C. F. Ong, M. S. DeMers, A. Rajagopal, M. Millard, S. R. Hamner, E. M. Arnold, J. R. Yong, S. K. Lakshmikanth, M. A. Sherman, J. P. Ku, S. L. Delp, OpenSim: Simulating musculoskeletal dynamics and neuromuscular control to study human and animal movement. *PLOS Comput. Biol.* **14**, e1006223 (2018).
32. G. S. Sawicki, O. N. Beck, I. Kang, A. J. Young, The exoskeleton expansion: Improving walking and running economy. *J. Neuroeng. Rehabil.* **17**, 25 (2020).
33. M. Serrao, G. Chini, M. Bergantino, D. Sarnari, C. Casali, C. Conte, A. Ranavolo, C. Marcotulli, M. Rinaldi, G. Coppola, F. Bini, F. Pierelli, F. Marinozzi, Dataset on gait patterns in degenerative neurological diseases. *Data Brief* **16**, 806–816 (2018).
34. T. Van Criekinge, W. Saeys, S. Truijen, L. Vereeck, L. H. Sloot, A. Hallemans, A full-body motion capture gait dataset of 138 able-bodied adults across the life span and 50 stroke survivors. *Sci. Data* **10**, 852 (2023).
35. P.-F. David, R.-C. David, J. C. Moreno, T. Diego, Human locomotion databases: A systematic review. *IEEE J. Biomed. Health Inform.* **28**, 1716–1729 (2024).
36. J.-Y. Zhu, T. Park, P. Isola, A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *2017 IEEE International Conference on Computer Vision (ICCV)* (IEEE, 2017), pp. 2242–2251.
37. S. An, A. Medda, M. N. Sawka, C. J. Hutto, M. L. Millard-Stafford, S. Appling, K. L. S. Richardson, O. T. Inan, AdaptNet: Human activity recognition via bilateral domain adaptation using semi-supervised deep translation networks. *IEEE Sens. J.* **21**, 20398–20411 (2021).
38. D. D. Molinaro, I. Kang, J. Camargo, M. C. Gombolay, A. J. Young, Subject-independent, biological hip moment estimation during multimodal overground ambulation using deep learning. *IEEE Trans. Med. Robot. Bionics* **4**, 219–229 (2022).
39. K. Scherpereel, D. Molinaro, O. Inan, M. Shepherd, A. Young, A human lower-limb biomechanics and wearable sensors dataset during cyclic and non-cyclic activities. *Sci. Data* **10**, 924 (2023).
40. J. Zhang, P. Fiers, K. A. Witte, R. W. Jackson, K. L. Poggensee, C. G. Atkeson, S. H. Collins, Human-in-the-loop optimization of exoskeleton assistance during walking. *Science* **356**, 1280–1284 (2017).
41. J. C. Selinger, J. M. Donelan, Estimating instantaneous energetic cost during non-steady-state gait. *J. Appl. Physiol.* **117**, 1406–1415 (2014).
42. Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc.: B (Methodol.)* **57**, 289–300 (1995).
43. K. L. Scherpereel, D. D. Molinaro, M. K. Shepherd, O. T. Inan, A. J. Young, Improving biological joint moment estimation during real-world tasks with EMG and instrumented insoles. *IEEE Trans. Biomed. Eng.* **71**, 2718–2727 (2024).
44. X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, S. Paul Smolley, "Least squares generative adversarial networks," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (IEEE, 2017), pp. 2794–2802.
45. Y. Taigman, A. Polyak, L. Wolf, Unsupervised cross-domain image generation. arXiv:1611.02200 [cs.CV] (2016).
46. H. Haresamudram, I. Essa, T. Plötz, Assessing the state of self-supervised human activity recognition using wearables. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **6**, 116 (2022).
47. F. J. Ordóñez, D. Roggen, Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition. *Sensors* **16**, 115 (2016).
48. O. Ronneberger, P. Fischer, T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, A. F. Frangi, Eds. (Lecture Notes in Computer Science, Springer International Publishing, 2015), pp. 234–241.
49. Y. Zhang, Z. Zhang, Y. Zhang, J. Bao, Y. Zhang, H. Deng, Human activity recognition based on motion sensor using U-Net. *IEEE Access* **7**, 75213–75226 (2019).
50. H. Ismail Fawaz, G. Forestier, J. Weber, L. Idoumghar, P.-A. Muller, Deep learning for time series classification: A review. *Data Min. Knowl. Disc.* **33**, 917–963 (2019).
51. S. Bai, J. Z. Kolter, V. Koltun, An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv:1803.01271 [cs.lG] (2018).
52. M. S. B. Hossain, Z. Guo, H. Choi, Estimation of lower extremity joint moments and 3D ground reaction forces using IMU sensors in multiple walking conditions: A deep learning approach. *IEEE J. Biomed. Health Inform.* **27**, 2829–2840 (2023).
53. J. Beil, I. Ehrenberger, C. Scherer, C. Mandery, T. Asfour, "Human motion classification based on multi-modal sensor data for lower limb exoskeletons," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (IEEE, 2018), pp. 5431–5436.

Acknowledgments: We thank R. Emadi for his tireless support collecting the model validation data. We also thank J. Leestma, K. Ghonasgi, and C. Nuesslein for reviewing the manuscript and D. Molinaro for being a sounding board for initial ideas. We acknowledge X, the Moonshot Factory, for providing the hardware to test these controllers. We also thank the research cyberinfrastructure resources and services provided by the Partnership for an Advanced Computing Environment (PACE) at the Georgia Institute of Technology, Atlanta, GA, USA. **Funding:** This work was supported by the National Science Foundation Graduate Research Fellowship no. DGE-2039655 (to K.L.S.) and the National Science Foundation Foundational Research in Robotics nos. 2233164 (to A.J.Y. and M.C.G.) and 2328051 and 2328050 (to A.J.Y. and M.K.S.). **Author contributions:** Conceptualization: K.L.S. and A.J.Y. Methodology: K.L.S., M.C.G., A.J.Y., and O.T.I. Software: K.L.S. Validation: K.L.S. Formal analysis: K.L.S. Investigation: K.L.S. and C.A.C. Resources: A.J.Y. Data curation: K.L.S. Visualization: K.L.S. Funding acquisition: K.L.S., M.C.G., M.K.S., and A.J.Y. Project administration: K.L.S. and A.J.Y. Supervision: K.L.S., A.J.Y., M.C.G., M.K.S., and O.T.I. Writing—original draft: K.L.S. and M.K.S. Writing—review and editing: K.L.S., M.K.S., A.J.Y., and M.C.G. **Competing interests:** K.L.S. and A.J.Y. are inventors on a US Patent application, patent pending, titled "Transfer learning of deep learning system and method for task agnostic wearable robot control and human monitoring," through the Georgia Institute of Technology, which describes some of the domain adaptation strategies used in this study. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper or the Supplementary Materials. The participant datasets for training are both open source: the data from Scherpereel *et al.* (*39*) can be found at https://doi.org/10.35090/gatech/70296 and the data from Molinaro *et al.* (*12*) can be found at https://doi.org/10.35090/gatech/75759. The new eight-participant dataset of time-synced exoskeleton data and human biomechanics and code for reproducing the main text figures is hosted in the SMARTech repository (https://doi.org/10.35090/gatech/79332).

Submitted 7 September 2024
Accepted 22 October 2025
Published 19 November 2025
10.1126/scirobotics.ads8652

Science Robotics

Deep domain adaptation eliminates costly data required for task-agnostic wearable robotic control

Keaton L. Scherpereel, Matthew C. Gombolay, Max K. Shepherd, Carlos A. Carrasquillo, Omer T. Inan, and Aaron J. Young

Sci. Robot. **10** (108), eads8652. DOI: 10.1126/scirobotics.ads8652

Science Robotics (ISSN 2470-9476) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science Robotics* is a registered trademark of AAAS.