

Homogeneous Dynamics Space for Heterogeneous Humans

Xinpeng Liu^{1,2}, Junxuan Liang¹, Chenshuo Zhang¹, Zixuan Cai³, Cewu Lu^{1,2*}, Yong-Lu Li^{1,2*}

¹Shanghai Jiao Tong University, ²Shanghai Innovation Institute, ³Soochow University
 xinpengliu0907@gmail.com, zxcai@stu.suda.edu.cn,
 {whitefork, zhangchenshuo, lucewu, yonglu.li}@sjtu.edu.cn

Abstract

Analyses of human motion kinematics have achieved tremendous advances. However, the production mechanism, known as human dynamics, is still uncovered. In this paper, we aim to push the understanding of data-driven human dynamics forward. We identify a major obstacle to this as the **heterogeneity** of existing human motion understanding efforts. Specifically, heterogeneity exists in not only the diverse kinematics representations and hierarchical dynamics representations but also the data from different domains, namely biomechanics and reinforcement learning. With an in-depth analysis of the existing heterogeneity, we propose to emphasize the **homogeneity** beneath heterogeneity: all of them represent the **homogeneous** fact of human motion, though from different perspectives. Given this, we propose **Homogeneous Dynamics Space (HDyS)** as a fundamental space for human dynamics by aggregating heterogeneous data and training a homogeneous latent space with inspiration from the inverse-forward dynamics procedure. HDyS achieves decent mapping between human kinematics and dynamics by leveraging the heterogeneous representations and datasets. We demonstrate the feasibility of HDyS with extensive experiments and applications. The project page is <https://foruck.github.io/HDyS>.

1. Introduction

Analyses on human motion have a wide range of applications, including animation [11, 60], healthcare [10, 55], and robotics [57, 70]. The computer vision community has made tremendous progress in understanding human kinematics with tasks like human reconstruction [24, 27, 34], action recognition [25, 26, 28, 29, 51], and motion generation [30, 32, 60]. However, the production mechanism hidden beneath human motion, known as human dynamics, is still limitedly explored.

In this paper, we aim to push the understanding of human

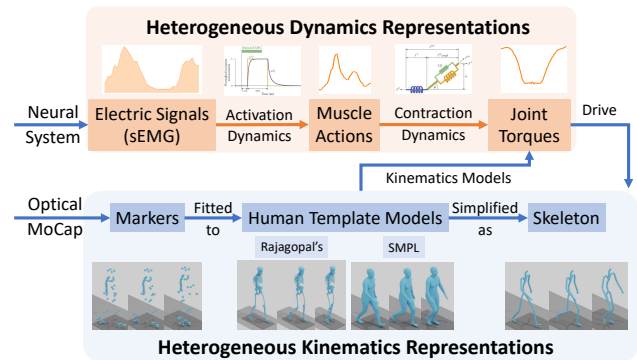


Figure 1. The representation heterogeneity exists for both kinematics and dynamics.

dynamics forward with data-driven methodologies. Specifically, we pursue to build a bidirectional mapping between human kinematics and dynamics. We identify the major obstacle to this as the heterogeneity of existing human motion understanding efforts, which could be two-fold.

First, **representation heterogeneity** exists for both kinematics and dynamics, as shown in Figure 1. A typical optical motion capture pipeline, which produces most existing available data for human kinematics [20, 40], involves tracking the *markers* placed on the surface of humans. Then, different human template models are fitted to the tracked markers, resulting in *parameters* of musculoskeletal models [52] for biomechanics uses or SMPL-like models [36] for general CV/CG uses. *Skeletons* with pre-defined kinematic trees are sometimes preferred as a simple and intuitive representation. Different datasets tend to include only representations preferred by their designed goal. Besides the kinematics representation heterogeneity, the dynamics representation heterogeneity exists more severely. Human dynamics function in a hierarchical manner. The closest dynamics hierarchy to the kinematics is the *joint torques*, which are further produced by muscle forces from *muscle actions* according to the muscle contraction dynamics. Furthermore, *electric signals* from the neural system activate the muscle, which surface electromyogra-

*Corresponding authors.

phy (sEMG) could detect. Effective conversions between these different representations are not always available. For example, for kinematics representations, the conversion between the musculoskeletal model and SMPL becomes available in very recent advances [20], which requires time-consuming optimization. For dynamics, the bi-directional mapping between joint torques, muscle actions, and electric signals is even more difficult.

Second, **domain heterogeneity** also exists. One crucial characteristic of human dynamics is the difficulty of non-intrusive measurement. In the biomechanics community, optimization-based methods are adopted to solve the Euler-Newton equation for joint torques [64] and muscle actions [53]. For electric signals, it is typically measured by sEMG sensors [3]. In the meantime, from the learning community, there emerge efforts [33] utilizing Reinforcement Learning (RL) to imitate existing motion captures with physically simulated humanoids and record the simulated human dynamics data in a fully synthetic manner. These different data sources introduce domain heterogeneity in three aspects. First, biomechanics data are more deeply rooted in reality with well-defined computation models and real human motions. Instead, RL data could diverge from real-world situations with unnatural motion imitations and inaccurate physics simulations, known as the sim2real domain gap. Second, biomechanics data, due to the optimization-based manner, is sensitive to variations and usually limited to strictly controlled laboratory setups with relatively simple motions, resulting in limited motion domain coverage. In contrast, RL data cover a wide range of motion with more diversity. Third, biomechanics data and RL data tend to adopt heterogeneous kinematics representations with different kinematic trees, making it hard to transfer directly from one domain to another.

Despite the heterogeneity, we emphasize the homogeneous essence behind it: all of them represent the homogeneous fact of human motion, each from a different perspective. Given this, we demonstrate the feasibility of unifying these heterogeneous human representations and exploiting the underlying homogeneous knowledge of human dynamics. Low-level Cartesian kinematics representations, like markers and joints, are less heterogeneous than representations like joint angles. Also, though joint torques, muscle actions, and electric signals are not directly transferable to each other, they could share similar motor knowledge like coordination. To this end, we propose **Homogeneous Dynamics Space** for heterogeneous humans (HDyS).

To achieve this, we aggregate human dynamics data from both RL [33] and biomechanics [3, 53, 65] with different kinematics and dynamics representations, covering human dynamics hierarchies, including joint torques, muscle actions, and electric signals. Then, HDyS is designed as an aggregation of multiple auto-encoders corresponding to the

inverse-forward dynamics procedure. We supervise HDyS with reconstruction and alignment losses. Our proposed HDyS has two major merits. First, by unifying heterogeneous human representation in the same latent space, it generalizes across different representations seamlessly while taking advantage of each. Second, by aggregating large-scale heterogeneous motion data, it is empowered with general human motor knowledge from a wide span of motion, functioning as a reusable knowledge source for downstream dynamics-related applications. To demonstrate the efficacy of HDyS, we first evaluate it on human inverse dynamics. Then, we showcase how HDyS could facilitate downstream dynamics-related applications like ground reaction force prediction, biomechanics human simulation, and physics-simulated character control.

To summarize, our contribution includes: 1) We analyzed the heterogeneity issue that hinders an in-depth understanding of human dynamics. 2) We highlighted the homogeneity beneath the heterogeneity and proposed a fundamental reusable space HDyS for human dynamics by unifying the heterogeneity. 3) We demonstrated the feasibility of digging homogeneity out from heterogeneity with extensive experiments and applications of HDyS.

2. Related Works

2.1. Human Dynamics

By human dynamics, we mean the production mechanism of human motion, which the biomechanics community has actively explored. To produce a certain motion, neural commands are sent to activate the muscles. After receiving the activation signals, muscles contract and produce muscle forces. Multiple muscle forces form the joint torques according to certain musculoskeletal geometry, and the joint torques drive the accelerations that accumulate into movements. Thus, understanding human dynamics typically involves two heterogeneous hierarchies: joint torques and muscle activations. However, both are hard to measure non-intrusively. In the literature, to obtain them, an optimization problem is typically introduced as

$$\min \|a\|, s.t. 0 \leq a \leq 1, \tau = A(q)F(a), \\ M(q)\ddot{q} + C(q, \dot{q}) + G(q) = J\lambda + \tau. \quad (1)$$

with the generalized human inertia matrix $M(q)$ w.r.t. generalized coordinate q , Coriolis and centrifugal forces $C(q, \dot{q})$, gravity $G(q)$, Jacobian matrix J mapping external forces λ to the generalized coordinates, and muscle activations a . The joint torques τ could be obtained by $\tau = A(q)F(a)$, where $A(q)$ maps muscle forces into joint torques and $F(a)$ maps a into muscle forces usually with the hill-type function [15, 74]. Mature software [4, 5, 63] were developed for this purpose. However, the optimization quality is tightly bonded to the precision of external

force measurement, which could be expensive. Therefore, the applications are mostly limited to simple motions like gaits in laboratory settings. Some efforts exploited wearable devices [22, 23] for more general applications. In addition, the optimization is deeply coupled with the adopted human models [36, 52], which vary with the application preferences. Fitting raw motion capture data to a specific human model could be unstable and time-consuming. The conversion between different human models is also non-trivial even with recent advances [20]. This way, transferring dynamics from one human model to another is limitedly explored. Besides torques and muscle actions that are typically obtained via optimization, Surface Electromyography (sEMG) is adopted as an indirect representation of human dynamics which could be directly measured. Efforts were made to build the accurate mapping from sEMG to muscle actions [19, 66], joint torques [54, 79], and human poses [35, 55]. Though progress was made, sEMG patterns might suffer from noises and vary drastically among subjects [61], hindering the generalization.

2.2. Learning-based Human Dynamics

Joint Torques. Early efforts were made on ML-based joint torque analysis for certain human body parts [17, 43, 69]. Lv *et al.* [39] developed a Gaussian mixture framework for whole-body joint torque estimation. Other architectures like k-nearest neighbor-based regression [75, 77], random forests [76], and neural networks [78] were also adopted for the estimation of joint torques. However, most of these efforts suffered from limited data scale, which hampered learning methods from exploiting their full potential. The recent emergence of AddBiomechanics [65], which aggregated multiple biomechanics datasets, considerably boosted the data scale. However, most of the collected sequences contained only gaits with limited diversity. Another line of work adopted reinforcement learning to simulate motion in physics simulators, and the joint torques could be obtained in simulation. Though generalizability was limited at first [1, 49, 50, 67, 68], emerging efforts [37, 38, 73] managed to replicate a wide span of motion in simulators. The paradigm is further incorporated into recent MoCap systems for simultaneous estimation of motion and the joint torques [8, 12, 16, 62, 72, 80]. However, due to the involvement of simulators, the sim2real gap exists. The learned dynamics could be restricted to certain simulators or human models, which might diverge from real humans.

Muscle Actions. sEMG is usually adopted as a proxy measurement of muscle actions and has been extensively studied in biomechanics [3, 6, 7, 14, 41, 44, 48]. There have been efforts delving into predicting sEMG signals given either joint torques [31, 56, 58], goniometers [59], motion captures [18, 45, 71], videos [3, 48], or point clouds [46]. However, each effort could be small in data scale, mo-

tion variations, and muscle coverage. Unifying these efforts could be hindered by the heterogeneous settings and data formats adopted, thus under-covered. Recently, musculoskeletal human simulation has gained attention. Multiple efficient simulators were developed [2, 9, 81], opening the potential for more accessible muscle action analysis. Specifically, MinT [53] was proposed by simulating motion sequences from AMASS [40] in OpenSim [5] and attaching simulated muscle actions to the raw motion sequence, enabling muscle action learning in scale.

3. Method

We introduce the proposed Homogeneous Dynamics Space (HDyS). We first cover the involved kinematics and dynamics representations in Section 3.1, 3.2. Then, the model architecture and designed losses are introduced in Section 3.3.

3.1. Kinematics Representations

As shown in Fig. 2, we adopt four types of kinematics representations: Cartesian representations of markers and skeleton key-points, joint angles of Rajagopal’s model [52] from the biomechanics community, and SMPL [36] which is widely adopted for CV and CG applications.

Markers, placed on the human body surfaces, are typically the raw data for optical motion capture. In practice, most other representations are calculated by fitting certain human prior models to the marker observations, making it easy to obtain for heterogeneous datasets and representations. Thus, markers are expected to be a generalizable representation across heterogeneous datasets, though they could also be rather low-level and thus hard to learn. We define the marker representation at timestamp t as $x_m^t = (m^t, \dot{m}^t, \ddot{m}^t) \in \mathcal{R}^{N_m \times 9}$, which is composed of marker Cartesian coordinates m^t , finite-differentiated velocities \dot{m}^t and finite-differentiated accelerations \ddot{m}^t with N_m markers.

Skeletal Keyoints, compared to markers, are less generalizable due to their reliance on pre-defined kinematic trees. However, with its Cartesian coordinates, common sense of human topology, and easy access, the gap between different kinematic trees can be mitigated. We define the joint representation at timestamp t as $x_k^t = (k^t, \dot{k}^t, \ddot{k}^t) \in \mathcal{R}^{N_k \times 9}$, which is composed of joint coordinates k^t , finite-differentiated velocities \dot{k}^t and finite-differentiated accelerations \ddot{k}^t with N_k skeletal joints.

Joint Angles are preferred for clinical analyses which more faithfully preserves the biomechanics information of human kinematics. We adopt the Rajagopal’s model [52] used in AddBiomechanics [65], and define the joint angle representation at timestamp t as $x_a^t = (a^t, \dot{a}^t, \ddot{a}^t) \in \mathcal{R}^{3N_j}$, containing joint angles a^t , finite-differentiated velocities \dot{a}^t and accelerations \ddot{a}^t with N_j joints in Rajagopal’s model (only 23 lower-body joints are used).

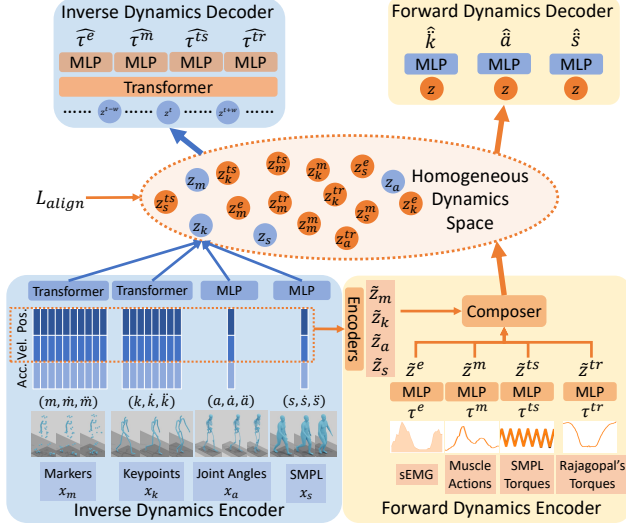


Figure 2. The overall architecture of HDyS.

SMPL [36] is a human parameter model widely adopted in the CV/CG community. We define the SMPL representation at timestamp t as $x_s^t = (s^t, \dot{s}^t, \ddot{s}^t) \in \mathcal{R}^{3N_j}$, which is composed of SMPL parameters s^t , finite-differentiated velocities \dot{s}^t and finite-differentiated accelerations \ddot{s}^t with N_j joints of SMPL (75 used, 3 translational joints at root and 72 revolving joints).

3.2. Dynamics Representations

We adopt three different types of dynamic representations: joint torques, muscle actions, and sEMGs.

Joint Torques are the net torques exerted at each joint to drive the motion, represented as $\tau_t \in \mathcal{R}^{N_t}$. N_t is coupled with the number of joints N_j as $N_t = N_j - 6$, with the 6-DoF free root joint unactuated. For clarity, we denote the joint torques corresponding to Rajagopal's model as τ_{tr} and those corresponding to SMPL as τ_{ts} .

Muscle Actions represent the activation level of each muscle, resulting in the muscle forces that produce joint torques. We denote muscle actions as $\tau_m \in \mathcal{R}^{N_m}$ with N_m muscles. **sEMG** detects the electric potential generated on the body surface by activated muscle cells, which is closely related to muscle actions. We represent sEMGs at timestamp t as $\tau_e^t \in \mathcal{R}^{N_e}$ with N_e sEMG channels.

3.3. Homogeneous Dynamics Space

With the heterogeneous representations, we construct our homogeneous dynamics space (HDyS) as an aggregation of multiple auto-encoders corresponding to the inverse-forward dynamics procedure. Then, *reconstruction* and *alignment* losses are adopted for supervision. For simplicity, all superscripts t are omitted.

Inverse-Dynamics Auto-Encoder (IDAE) encodes the kinematics into the latent space and decodes the dynamics

from the latent. For markers x_m and skeletal key points x_k , we adopt three-layer transformer encoders with no positional embedding to encode them into latent $z_m, z_k \in \mathcal{R}^d$ of dimension d . This enables the encoding of an arbitrary number of markers/points. For joint angles x_a and SMPL parameters x_s , we adopt simple three-layer MLPs to encode them into latent $z_a, z_s \in \mathcal{R}^d$. The decoder is designed as a shared transformer followed by separate MLP regression heads. The transformer takes latent vectors from consecutive frames as input, outputting *per-frame* latent vectors refined with temporal contexts. Then, separate MLP decoders decode the per-frame dynamics as $\hat{\tau}_{tr}, \hat{\tau}_{ts}, \hat{\tau}_m, \hat{\tau}_e$.

Forward-Dynamics Auto-Encoder (FDAE) encodes dynamics and the kinematics except for accelerations (denoted as \tilde{x}) into the latent space and decodes the accelerations. Similar to IDAE, markers \tilde{x}_m and skeletal key-points \tilde{x}_k are encoded by transformers as \tilde{z}_m, \tilde{z}_k , while joint angles \tilde{x}_a and SMPL parameters \tilde{x}_s are encoded by MLPs as \tilde{z}_a, \tilde{z}_p . Then, joint torques τ_{tr}, τ_{ts} , muscle activations τ_m , and sEMGs τ_e are also encoded by separate MLP encoders as dynamics latent vectors $\tilde{z}^{tr}, \tilde{z}^{ts}, \tilde{z}^m, \tilde{z}^e$. $\tilde{z}_m, \tilde{z}_k, \tilde{z}_a, \tilde{z}_p$ is then *concatenated* with available dynamics latent vectors, fed into a shared MLP composer, resulted in latent vectors z , with subscripts representing the kinematics and superscripts representing the dynamics. Although arbitrary combinations are feasible, in practice, only one dynamic latent vector typically exists due to the data limitations. Finally, separate MLP decoders decode the latent vectors $z_{\{kin\}}^{\{dyn\}}$ into skeletal key point accelerations \ddot{k} , SMPL accelerations \ddot{s} , and joint angle accelerations \ddot{a} , where $kin \in \{m, k, s, a\}$, $dyn \in \{m, e, ts, tr\}$. The marker accelerations are not predicted since their variable sizes could bring unnecessary complexity.

Loss Terms. To train HDyS, we adopt reconstruction losses and alignment losses. For reconstruction losses, we simply calculate the L1 loss for τ and \ddot{a} as

$$L_{recon} = \|\tau - \hat{\tau}\|_1 + \|\ddot{a} - \hat{\ddot{a}}\|. \quad (2)$$

For alignment losses, we align the latent vectors z of the same frame and separate z s of different frames using InfoNCE [47]. Given a batch of B frames, the obtained latent vectors is denoted as $Z = \{z_{\{kin\}}, z_{\{kin\}}^{\{dyn\}}\}$, with $kin \in \{m, k, s, a\}$, $dyn \in \{m, e, ts, tr\}$. The loss is

$$L_{align} = \sum_{z_1, z_2 \in Z} \sum_{i=1}^B -\log\left(\frac{\exp(\langle z_1^{(i)}, z_2^{(i)} \rangle)}{\sum_{j=1}^B \exp(\langle z_1^{(i)}, z_2^{(j)} \rangle)}\right), \quad (3)$$

where $(i), (j)$ indicate the batch index. The overall loss is as $\mathcal{L} = \alpha_1 L_{recon} + \alpha_2 L_{align}$ with coefficients α_1, α_2 .

Table 1. Quantitative results of HDyS with ablation studies. For HDyS, the results of *averaged* predictions and the *best* prediction among all representations are reported.

Methods	ImDy mPJE↓ avg/bst	AddBiomechanics mPJE↓ avg/bst	MinT RMSE↓ avg/bst	PCC↑ avg/bst	MiA RMSE↓ avg/bst	PCC↑ avg/bst
ImDyS [33]	0.6300	0.1626	-	-	-	-
MiA[3]	-	-	-	-	13.3	-
HDyS	0.5765/0.4674	0.1189/0.1243	0.0614/0.0615	0.7420/0.7402	11.8/11.6	0.7232/0.7261
ImDy-only HDyS	0.6854/0.5403	-	-	-	-	-
AddBiomechanics-only HDyS	-	0.1695/0.1691	-	-	-	-
MinT-only HDyS	-	-	0.0637/0.0640	0.7179/0.7127	-	-
MiA-only HDyS	-	-	-	-	13.6/13.5	0.6557/0.6421
HDyS w/o ImDy	-	0.1214/0.1386	0.0608/0.0617	0.7470/0.7417	15.8/15.6	0.6523/0.6497
HDyS w/o AddBiomechanics	0.5787/0.4742	-	0.0616/ 0.0614	0.7408/0.7386	17.1/17.0	0.5769/0.5638
HDyS w/o MinT	0.5730/0.4681	0.1197/0.1296	-	-	16.9/16.8	0.5213/0.5313
HDyS w/o MiA	0.5890/0.4788	0.1200/0.1284	0.0616/0.0618	0.7395/0.7375	-	-
HDyS w/o AMASS	0.5797/0.4786	0.1217/0.1319	0.0613/0.0617	0.7419/0.7380	14.7/14.9	0.5704/0.5632
HDyS w/o L_{align}	0.6575/0.5019	0.1270/0.1329	0.0626/0.0630	0.7318/0.7238	13.7/13.4	0.6464/0.6402
HDyS w/o FDAE	0.5776/0.4849	0.1198/ 0.1261	0.0617/0.0617	0.7388/0.7375	13.6/ 13.3	0.6517/ 0.6699
HDyS-32D	0.7390/0.6450	0.1401/0.1420	0.0650/0.0648	0.6980/0.7010	16.7/16.3	0.5524/0.5614
HDyS-64D	0.6505/0.5410	0.1295/0.1354	0.0627/0.0629	0.7272/0.7254	15.1/14.5	0.6034/0.6187

4. Experiments

4.1. Datasets

AddBiomechanics [64] contains over 50 hours of human motion data with joint torques from Nimble [63] simulation. We follow the setting in [64] with the armless part of AddBiomechanics. Markers, key points, and joint angles are used as kinematics representations, and joint torques corresponding to [52] represent the dynamics.

Muscles in Times (MinT) [53] simulates part of AMASS in OpenSim to obtain the actions of 402 muscles. We randomly split them into 906 sequences for training and 227 sequences for testing. Markers, key points, and SMPL parameters are used as kinematics representations, and muscle actions are the dynamics representations.

Muscles in Act (MiA) [3] consists of 12.8 hours of human exercise motion reconstructed from videos with VIBE [21] and sEMG data of 8 muscles. Following [3], we split them into 19,563 training sequences and 3,053 testing sequences. Markers and key points are kinematics representations, and sEMGs are dynamics representations.

ImDy [33] adopted PHC [38] to imitate motion sequences from AMASS [40] and KIT [42], resulting in over 150 hours of human dynamics data, with 27,501 sequences for training and 3,055 sequences for evaluation. Markers, key points, and SMPL are kinematics representations, and SMPL torques are dynamics representations.

AMASS [40] contains over 11,000 motion sequences represented in SMPL [36]. Though not paired with human dynamics, we included it as an extra knowledge base for training. Markers, key points, and SMPL parameters are used as kinematics representations.

4.2. Implementation Details

We adopt a compact design for HDyS with a latent dimension of 128. For joint angle and SMPL parameters, three-layer MLPs with hidden unit sizes of 256 and 128 are adopted as encoders, projecting the input to the 128-D latent space. For markers and key-points, three-layer transformer encoders with 2 heads are adopted as encoders. The transformer in the ID decoder is of 4 layers, 4 heads, and a dimension of 128. All MLPs in decoders are two-layer, with a hidden size of 32 for Rajagopal’s torque, sEMG, joint angle accelerations, and key-point accelerations, and a hidden size of 64 for SMPL torque, SMPL accelerations, and muscle actions. For training, we adopt an AdamW optimizer with a learning rate of 1e-3 and a batch size of 9,600 frames for 1,000 epochs. Loss weights are set as $\alpha_1 = 0.01, \alpha_2 = 0.05$. Also, during training, a balanced sampling strategy is adopted to minimize the scale influence of different datasets. That is, we randomly sample 3,000 sequences per dataset for each training epoch. In this way, the comparison is fair in terms of the number of seen samples under different dataset settings. All experiments are conducted on a single NVIDIA Titan Xp GPU.

4.3. Results on Inverse Dynamics

Metric. We report mPJE and RMSE as

$$mPJE_{\tau} = \frac{1}{J} \sum_{j=1}^J |\tau_j - \hat{\tau}_j|_2, RMSE_{\tau} = |\tau - \hat{\tau}|_2. \quad (4)$$

PCC is also reported due to its invariance against scale and offset, which helps evaluate muscle actions and sEMG prediction. PCC is computed as

$$PCC_{\tau} = \frac{cov(\tau, \hat{\tau})}{\sigma_{\tau} \sigma_{\hat{\tau}}}, \quad (5)$$

where $cov(\tau, \hat{\tau})$ is the covariance of τ , $\hat{\tau}$, and σ indicates the standard deviation. For ImDy and AddBiomechanics, we report mPJE normalized by body weight. For MinT and MiA, RMSE and PCC are reported.

4.3.1 Quantitative results

Quantitative results are shown in Tab. 1, where HDyS is reported with the averaged predictions from different input kinematics representations and the best predictions among all representations. Compared to previous methods [3, 33], the proposed HDyS provides superior performances with substantial improvements on all datasets. We further analyze the performance of HDyS with three questions.

How do the heterogeneous datasets contribute to HDyS?

We compare HDyS with its single-dataset variants and drop-one-out variants in Tab. 1. HDyS outperforms the corresponding single-dataset variants on all datasets, validating the existence of homogeneous human dynamics knowledge behind these heterogeneous datasets. According to the drop-one-out results, it is noticeable that datasets with similar dynamics representations are more cooperative. That is, muscle-related datasets (MiA and MinT) tend to benefit more from each other and less from torque-related datasets (AddBiomechanics and ImDy), and vice versa. Moreover, mutually harmful effects could be observed for ImDy and MinT. These are consistent with the gaps between these datasets: the sim2real gap from ImDy to others and the torque-to-muscle gap from AddBiomechanics and ImDy to MiA and MinT. AMASS, though not paired with dynamics information, is shown to be beneficial for inverse dynamics with its diverse and high-quality kinematics.

The heterogeneous datasets introduce both increased data scale and heterogeneous knowledge. Given this, we try to investigate further the source of the improvement in Tab. 1. Trying to decompose the contributions of scale and heterogeneity, we compared the performance of the following three models on a single target test set in Tab. 2. *HDyS-Single-50* denotes single-dataset HDyS with 50% of the data from the target dataset. *HDyS-50/50* denotes HDyS using 50% of the data from the target dataset and 50% of the data from the other datasets, maintaining the same data scale as the target dataset. *HDyS-Single* represents the corresponding single-dataset variants of HDyS in Tab. 1. The performance of the best representation is reported on AddBiomechanics and MiA, considering their suitable scale and realistic nature. As shown, HDyS-50/50 consistently outperforms HDyS-Single-50, verifying the existence of homogeneous knowledge in heterogeneous data. Moreover, on AddBiomechanics, HDyS-50/50 even outperforms HDyS-Single, indicating that the diversity from heterogeneous data could sometimes benefit more than simply increasing seemingly homogeneous data.

Table 2. Results on scale-heterogeneity decomposition.

Dataset	HDyS-Single-50	HDyS-50/50	HDyS-Single
AddBiomechanics <i>mPJE</i> ↓	0.1707	0.1284	0.1695
MiA <i>RMSE</i> ↓	16.2	14.5	13.5

Table 3. Results of different kinematic representations in HDyS.

Methods	ImDy	AddBiomechanics	MinT	MiA
	mPJE↓	mPJE↓	RMSE↓ PCC↑	RMSE↓ PCC↑
HDyS-marker	0.8163	0.1455	0.0646 0.6968	12.3 0.7088
HDyS-keypoint	0.8084	0.1324	0.0640 0.7103	11.6 0.7261
HDyS-angle	-	0.1243	-	-
HDyS-SMPL	0.4674	-	0.0615 0.7402	-

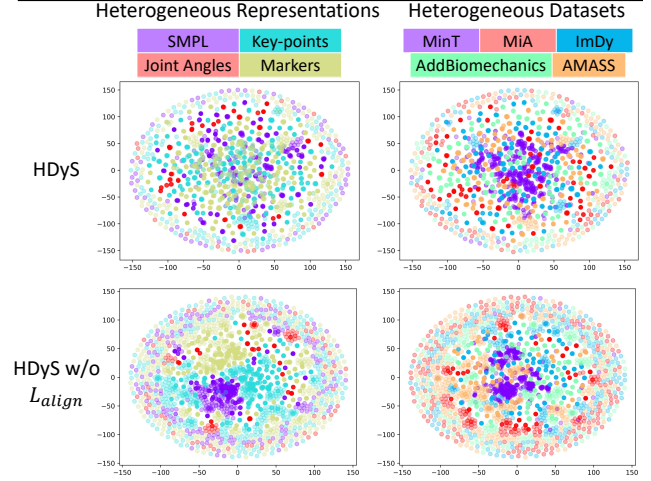


Figure 3. Feature visualization for HDyS w/ and w/o L_{align} .

How do the heterogeneous representations contribute to HDyS?

We report the results from individual kinematics representations in Tab. 3. Typically, the representations in joint space like SMPL and joint angles produce more accurate predictions compared to Cartesian representations with considerable performance gaps, which explains the performance drop when taking the average for ImDy and MiA. The markers are the worst-performed representation of all.

In Tab. 1, it is noticeable that MinT and AddBiomechanics take advantage of averaging predictions from different representations. For AddBiomechanics, this might be because the best-performing representation is joint angles, which are not available in other datasets. Therefore, the averaging could effectively aggregate knowledge from other datasets with other representations. For MinT, it is observable that different representations produce close results, preventing the ensemble from being held back by markers.

Furthermore, we examine the alignment of heterogeneous representations by removing the alignment loss L_{align} in Tab. 1. Considerable performance drops could be observed. The drop is consistent for both best and average performances, indicating the cross-modality alignment manages to aggregate knowledge from heterogeneous representations. We also visualize HDyS features w/ and w/o L_{align} in Fig. 3. As shown, L_{align} prevents the HDyS feature from being dominated by the representation hetero-

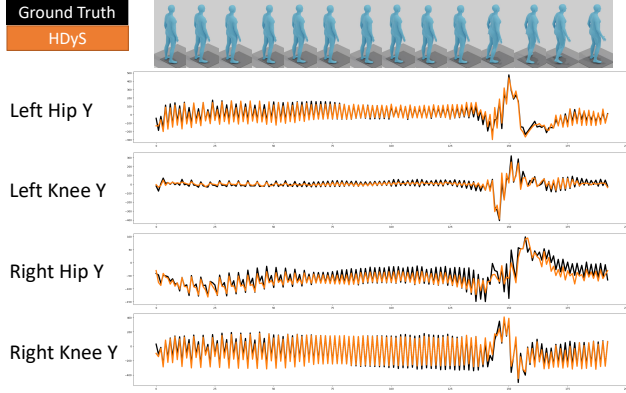


Figure 4. Inverse dynamics visualization on ImDy.

generality, encouraging the learning of homogeneous knowledge. Interestingly, though not intended to do so, L_{align} also helps to mitigate the gap between datasets, alleviating the domain heterogeneity.

How does the model design contribute to HDyS? We first evaluate the FDAE in Tab. 1. A substantial performance drop could be observed, demonstrating the benefit of informing HDyS with the forward dynamics procedure. Moreover, we also report HDyS with different dimensions in Tab. 1. Though HDyS’s performance degrades with lower dimensions, it could still outperform some single-dataset variants with higher dimensions, validating the efficacy of heterogeneous datasets again.

4.3.2 Qualitative results

We visualize the inverse dynamics results of different datasets. As shown in Fig. 4-5, the joint torques could be faithfully reconstructed for different actions from either synthetic and jittering ImDy or realistic AddBiomechanics with the same HDyS. For the muscle actions shown in Fig. 6, HDyS produces reasonable estimations for lower-body muscles when walking. Moreover, as in the right half, though the lower-body kinematics are less significant, HDyS could capture the ankle muscle actions with high fidelity. The predictions are also coherent with GT for upper-body muscles like the internal oblique and levator scapulae. More details are in the appendix.

4.4. Results on Downstream Tasks

4.4.1 Ground Reaction Force Estimation

We evaluate the effectiveness of HDyS on GRF estimation with GroundLink [13], which contains 1.5-hour motion with GRF recordings. We finetune HDyS with subjects 1-6 and evaluate it on subject 7. mPJE for GRF at both feet normalized by body weight is reported following Eq. 4. GroundLinkNet [13] and HDyS trained on GroundLink from scratch are also compared. As shown in Tab. 4, the

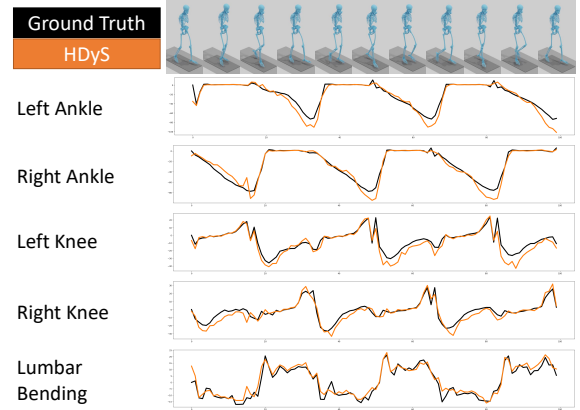


Figure 5. Inverse dynamics visualization on AddBiomechanics. Table 4. Results on GroundLink.

Methods	GroundLinkNet	GroundLink-only HDyS	HDyS
L-Foot $mPJE \downarrow$	0.0711	0.0584	0.0514
R-Foot $mPJE \downarrow$	0.0912	0.0751	0.0694

Table 5. Biomechanical human simulation results reported in per-frame MSE.

k	1	2	3	4	5
Optimized	0.00063	0.1909	1.8306	2.0106	2.6027
HDyS	0.00061	0.1860	1.7118	1.8651	2.2233

finetuned HDyS outperforms its counterparts, indicating the efficacy of the aggregated homogeneous knowledge. Also, the GroundLink-only HDyS outperforms GroundLinkNet, reflecting the feasibility of unifying heterogeneous representations for better dynamics knowledge.

4.4.2 Biomechanical Human Simulation

HDyS could also be adopted for biomechanical human simulation. We start with the armless Rajagopal’s model [52] in Nimble [63]. Given a motion sequence, we first adopt HDyS to estimate the joint torques. Then, we use the predicted torques to reproduce the motion in Nimble. Starting from the current state, we feed the predicted torques for k steps and compare the simulated joint angles \hat{q} with the real joint angles q . The simulation is performed at 90FPS. We report the per-frame MSE of joint angles.

Results. We demonstrated the results in Tab. 5 and Fig. 7. The results with optimized torques are also reported as a reference. Surprisingly, for different k , simulation with HDyS is superior with better stability. However, with k increasing, MSE also increases noticeably, indicating the accumulation of drifting errors. Further enhancing the forward-dynamics compatibility of HDyS would be a promising goal.

4.4.3 Physical Character Control

We adopt HDyS for physical character control following the setting of PHC [38], with 140 testing sequences and the rest

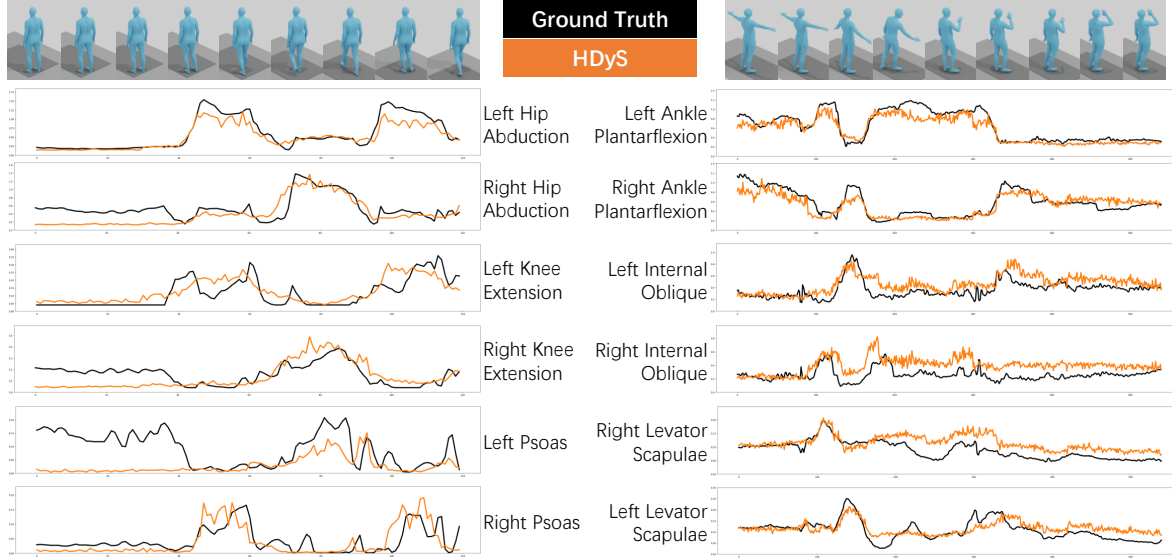


Figure 6. Inverse dynamics results on MinT.

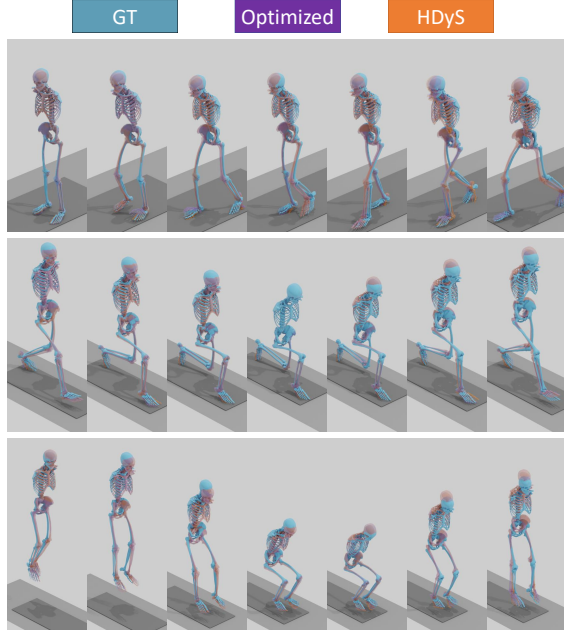


Figure 7. Biomechanical human simulation visualization.

Table 6. Results of HDyS for character control.

Methods	mPJPE _g ↓	mPJPE _l ↓	E _{vel} ↓	E _{acc} ↓
Baseline	74.514	49.029	11.562	14.243
Baseline + HDyS	72.692	47.920	11.305	13.757

for training. We train the baseline as a primitive in PHC with a batch size of 700 for 10k steps. Then, the HDyS latents of key points are inserted as extra observations. We report the global and local mPJPE and the errors of velocity and acceleration in Tab. 6. HDyS manages to improve its performance, validating its efficacy.

5. Discussion

Despite the impressive results of HDyS, it could be improved. First, it is noticeable that HDyS performs better for the lower body than the higher body in Fig. 6. This might be due to the imbalanced focus on lower body dynamics in data adopted by HDyS. For all datasets, data on lower-body dynamics like gaits are *dominating*. For AddBiomechanics, we only adopted its armless part. Enhancing HDyS with more upper-body dynamics would be helpful. Second, for muscle actions, HDyS could omit minor changes, and the magnitudes could sometimes diverge from the real. Mitigating it with more curated models and loss terms is desirable. Third, as a first step toward homogeneous human dynamics learning, HDyS is primarily instantiated with five initial datasets and an intuitive model. Scaling HDyS up with more datasets and more human priors would be a meaningful goal. Finally, we demonstrate the potential of HDyS with some simple applications in Sec. 4.4, while more sophisticated use cases for musculoskeletal human simulation, humanoid control, and human-robot transfer learning would be promising as future works.

6. Conclusion

We analyzed the heterogeneity issue existing for human dynamics learning and highlighted the homogeneity beneath it. To fully exploit the homogeneity, we proposed HDyS as a homogeneous human dynamics space. Extensive experiments were conducted to validate the feasibility of digging homogeneity out from heterogeneity for human dynamics with detailed analyses of the contribution of heterogeneous components. We further demonstrated the potential of HDyS for downstream applications. We believe HDyS could shed new light on human dynamics understanding.

Acknowledgements

This work is supported in part by the National Natural Science Foundation of China under Grant No.62306175, CCF-Tencent Rhino-Bird Open Research Fund.

References

- [1] Kevin Bergamin, Simon Clavet, Daniel Holden, and James Richard Forbes. Drecon: data-driven responsive control of physics-based characters. *ACM Transactions On Graphics (TOG)*, 38(6):1–11, 2019. **3**
- [2] Vittorio Caggiano, Huawei Wang, Guillaume Durandau, Massimo Sartori, and Vikash Kumar. Myosuite: A contact-rich simulation suite for musculoskeletal motor control. In *Learning for Dynamics and Control Conference*, pages 492–507. PMLR, 2022. **3**
- [3] Mia Chiquier and Carl Vondrick. Muscles in action. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 22091–22101, 2023. **2, 3, 5, 6**
- [4] Michael Damsgaard, John Rasmussen, Søren Tørholm Christensen, Egidijus Surma, and Mark De Zee. Analysis of musculoskeletal systems in the anybody modeling system. *Simulation Modelling Practice and Theory*, 14(8): 1100–1111, 2006. **2**
- [5] Scott L Delp, Frank C Anderson, Allison S Arnold, Peter Loan, Ayman Habib, Chand T John, Eran Guendelman, and Darryl G Thelen. Opensim: open-source software to create and analyze dynamic simulations of movement. *IEEE transactions on biomedical engineering*, 54(11):1940–1950, 2007. **2, 3**
- [6] Benedikt Feldotto, Cristian Soare, Alois Knoll, Piyanee Sriya, Sarah Astill, Marc de Kamps, and Samit Chakrabarty. Evaluating muscle synergies with emg data and physics simulation in the neurorobotics platform. *Frontiers in Neuro-robotics*, 16:856797, 2022. **3**
- [7] Mariusz P Furmanek, Madhur Mangalam, Mathew Yarossi, Kyle Lockwood, and Eugene Tunik. A kinematic and emg dataset of online adjustment of reach-to-grasp movements to visual perturbations. *Scientific data*, 9(1):23, 2022. **3**
- [8] E. Gärtner, M. Andriluka, E. Coumans, and C. Sminchisescu. Differentiable dynamics for articulated 3d human motion reconstruction. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 13180–13190, Los Alamitos, CA, USA, 2022. IEEE Computer Society. **3**
- [9] Thomas Geijtenbeek. The Hyfydy simulation software, 2021. <https://hyfydy.com>. **3**
- [10] Daniel FN Gordon, Andreas Christou, Theodoros Stouraitis, Michael Gienger, and Sethu Vijayakumar. Learning personalised human sit-to-stand motion strategies via inverse musculoskeletal optimal control. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 10497–10503. IEEE, 2023. **1**
- [11] Chuan Guo, Shihao Zou, Xinxin Zuo, Sen Wang, Wei Ji, Xingyu Li, and Li Cheng. Generating diverse and natural 3d human motions from text. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5142–5151, Los Alamitos, CA, USA, 2022. IEEE Computer Society. **1**
- [12] Erik Gärtner, Mykhaylo Andriluka, Hongyi Xu, and Cristian Sminchisescu. Trajectory optimization for physics-based reconstruction of 3d human pose from monocular video. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 13096–13105, Los Alamitos, CA, USA, 2022. IEEE Computer Society. **3**
- [13] Xingjian Han, Ben Senderling, Stanley To, Deepak Kumar, Emily Whiting, and Jun Saito. Groundlink: A dataset unifying human body movement and ground reaction dynamics. In *SIGGRAPH Asia 2023 Conference Papers*, New York, NY, USA, 2023. Association for Computing Machinery. **7**
- [14] Óscar G Hernández, Jose M Lopez-Castellanos, Carlos A Jara, Gabriel J Garcia, Andres Ubeda, Vicente Morell-Gimenez, and Francisco Gomez-Donoso. A kinematic, imaging and electromyography dataset for human muscular manipulability index prediction. *Scientific Data*, 10(1):132, 2023. **3**
- [15] Archibald Vivian Hill. The heat of shortening and the dynamic constants of muscle. *Proceedings of the Royal Society of London. Series B-Biological Sciences*, 126(843):136–195, 1938. **2**
- [16] B. Huang, L. Pan, Y. Yang, J. Ju, and Y. Wang. Neural moccon: Neural motion control for physically plausible human motion capture. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6407–6416, Los Alamitos, CA, USA, 2022. IEEE Computer Society. **3**
- [17] Leif Johnson and Dana H. Ballard. Efficient codes for inverse dynamics during walking. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, pages 343–349, Québec City, Québec, Canada, 2014. AAAI Press. **3**
- [18] Lise A Johnson and Andrew J Fuglevand. Evaluation of probabilistic methods to predict muscle activity: implications for neuroprosthetics. *Journal of neural engineering*, 6(5):055008, 2009. **3**
- [19] Kimoon Kang, Kiwon Rhee, and Hyun-Chool Shin. Event detection of muscle activation using an electromyogram. *Applied Sciences*, 10(16):5593, 2020. **3**
- [20] Marilyn Keller, Keenon Werling, Soyong Shin, Scott Delp, Sergi Pujades, C Karen Liu, and Michael J Black. From skin to skeleton: Towards biomechanically accurate 3d digital humans. *ACM Transactions on Graphics (TOG)*, 42(6): 1–12, 2023. **1, 2, 3**
- [21] Muhammed Kocabas, Nikos Athanasiou, and Michael J Black. Vibe: Video inference for human body pose and shape estimation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5253–5263, 2020. **5**
- [22] Claudia Latella, Naveen Kuppaswamy, Francesco Romano, Silvio Traversaro, and Francesco Nori. Whole-body human inverse dynamics with distributed micro-accelerometers, gyros and force sensing. *Sensors*, 16(5), 2016. **3**
- [23] Claudia Latella, Silvio Traversaro, Diego Ferigo, Yeshasvi Tirupachuri, Lorenzo Rapetti, Francisco Javier An-

- drade Chavez, Francesco Nori, and Daniele Pucci. Simultaneous floating-base estimation of human kinematics and joint torques. *Sensors*, 19(12), 2019. 3
- [24] Jiefeng Li, Chao Xu, Zhicun Chen, Siyuan Bian, Lixin Yang, and Cewu Lu. Hybrik: A hybrid analytical-neural inverse kinematics solution for 3d human pose and shape estimation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3383–3393, 2021. 1
- [25] Sheng-Wei Li, Zi-Xiang Wei, Wei-Jie Chen, Yi-Hsin Yu, Chih-Yuan Yang, and Jane Yung-jen Hsu. Sa-dvae: Improving zero-shot skeleton-based action recognition by disentangled variational autoencoders. In *European Conference on Computer Vision*, pages 447–462. Springer, 2024. 1
- [26] Yong-Lu Li, Siyuan Zhou, Xijie Huang, Liang Xu, Ze Ma, Hao-Shu Fang, Yanfeng Wang, and Cewu Lu. Transferable interactiveness knowledge for human-object interaction detection. In *CVPR*, 2019. 1
- [27] Yong-Lu Li, Xinpeng Liu, Han Lu, Shiyi Wang, Junqi Liu, Jiefeng Li, and Cewu Lu. Detailed 2d-3d joint representation for human-object interaction. In *CVPR*, 2020. 1
- [28] Yong-Lu Li, Liang Xu, Xinpeng Liu, Xijie Huang, Yue Xu, Shiyi Wang, Hao-Shu Fang, Ze Ma, Mingyang Chen, and Cewu Lu. Pastanet: Toward human activity knowledge engine. In *CVPR*, 2020. 1
- [29] Yong-Lu Li, Xinpeng Liu, Xiaoqian Wu, Yizhuo Li, Zuoyu Qiu, Liang Xu, Yue Xu, Hao-Shu Fang, and Cewu Lu. Hake: A knowledge engine foundation for human activity understanding. *TPAMI*, 2022. 1
- [30] Yong-Lu Li, Xiaoqian Wu, Xinpeng Liu, Yiming Dou, Yikun Ji, Junyi Zhang, Yixing Li, Jingru Tan, Xudong Lu, and Cewu Lu. From isolated islands to pangea: Unifying semantic space for human action understanding. In *CVPR*, 2024. 1
- [31] Zhan Li, David Guiraud, and Mitsuhiro Hayashibe. Inverse estimation of multiple muscle activations from joint moment with muscle synergy extraction. *IEEE journal of biomedical and health informatics*, 19(1):64–73, 2014. 3
- [32] Xinpeng Liu, Haowen Hou, Yanchao Yang, Yong-Lu Li, and Cewu Lu. Revisit human-scene interaction via space occupancy. In *European Conference on Computer Vision*, pages 1–19. Springer, 2024. 1
- [33] Xinpeng Liu, Junxuan Liang, Zili Lin, Haowen Hou, Yong-Lu Li, and Cewu Lu. Imdy: Human inverse dynamics from imitated observations. In *The Thirteenth International Conference on Learning Representations*, 2025. 2, 5, 6
- [34] Xiaoyang Liu, Boran Wen, Xinpeng Liu, Zizheng Zhou, Hongwei Fan, Cewu Lu, Lizhuang Ma, Yulong Chen, and Yong-Lu Li. Interacted object grounding in spatio-temporal human-object interactions. In *AAAI*, 2025. 1
- [35] Yilin Liu, Shijia Zhang, and Mahanth Gowda. Neuropose: 3d hand pose tracking using emg wearables. In *Proceedings of the Web Conference 2021*, pages 1471–1482, 2021. 3
- [36] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. Smpl: a skinned multi-person linear model. In *ACM Transactions on Graphics*, New York, NY, USA, 2015. Association for Computing Machinery. 1, 3, 4, 5
- [37] Zhengyi Luo, Ryo Hachiuma, Ye Yuan, and Kris Kitani. Dynamics-regulated kinematic policy for egocentric pose estimation. *Advances in Neural Information Processing Systems*, 34:25019–25032, 2021. 3
- [38] Z. Luo, J. Cao, A. Winkler, K. Kitani, and W. Xu. Perpetual humanoid control for real-time simulated avatars. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 10861–10870, Los Alamitos, CA, USA, 2023. IEEE Computer Society. 3, 5, 7
- [39] Xiaolei Lv, Jinxiang Chai, and Shihong Xia. Data-driven inverse dynamics for human motion. *ACM Transactions on Graphics (TOG)*, 35(6):1–12, 2016. 3
- [40] N. Mahmood, N. Ghorbani, N. F. Troje, G. Pons-Moll, and M. Black. Amass: Archive of motion capture as surface shapes. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 5441–5450, Los Alamitos, CA, USA, 2019. IEEE Computer Society. 1, 3, 5
- [41] Nebojša Malešević, Alexander Olsson, Paulina Sager, Elin Andersson, Christian Cipriani, Marco Controzzi, Anders Björkman, and Christian Antfolk. A database of high-density surface electromyogram signals comprising 65 isometric hand gestures. *Scientific Data*, 8(1):63, 2021. 3
- [42] Christian Mandery, Ömer Terlemez, Martin Do, Nikolaus Vahrenkamp, and Tamim Asfour. Unifying representations and large-scale whole-body motion databases for studying human motion. *IEEE Transactions on Robotics*, 32(4):796–809, 2016. 5
- [43] Mykhailo Manukian, Serhii Bahdasariants, and Sergiy Yakovenko. Artificial physics engine for real-time inverse dynamics of arm and hand movement. *Plos one*, 18(12):e0295750, 2023. 3
- [44] Luís Moreira, Joana Figueiredo, Pedro Fonseca, João P Vilas-Boas, and Cristina P Santos. Lower limb kinematic, kinetic, and emg data from young healthy humans during walking at controlled speeds. *Scientific data*, 8(1):103, 2021. 3
- [45] Yoshihiko Nakamura, Katsu Yamane, Yusuke Fujita, and Ichiro Suzuki. Somatosensory computation for man-machine interface from motion-capture data and musculoskeletal human model. *IEEE Transactions on Robotics*, 21(1):58–66, 2005. 3
- [46] Hui Niu, Takahiro Ito, Damien Desclaux, Ko Ayusawa, Yusuke Yoshiyasu, Ryusuke Sagawa, and Eiichi Yoshida. Estimating muscle activity from the deformation of a sequential 3d point cloud. *Journal of Imaging*, 8(6):168, 2022. 3
- [47] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018. 4
- [48] Kunyu Peng, David Schneider, Alina Roitberg, Kailun Yang, Jiaming Zhang, Chen Deng, Kaiyu Zhang, M Saquib Sarfraz, and Rainer Stiefelhagen. Towards video-based activated muscle group estimation in the wild. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 4495–4504, 2024. 3
- [49] Xue Bin Peng, Ze Ma, Pieter Abbeel, Sergey Levine, and Angjoo Kanazawa. Amp: Adversarial motion priors for stylized physics-based character control. *ACM Transactions on Graphics (ToG)*, 40(4):1–20, 2021. 3

- [50] Xue Bin Peng, Yunrong Guo, Lina Halper, Sergey Levine, and Sanja Fidler. Ase: Large-scale reusable adversarial skill embeddings for physically simulated characters. *ACM Transactions On Graphics (TOG)*, 41(4):1–17, 2022. 3
- [51] A. R. Punnakal, A. Chandrasekaran, N. Athanasiou, A. Quiros-Ramirez, and M. J. Black. Babel: Bodies, action and behavior with english labels. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 722–731, Los Alamitos, CA, USA, 2021. IEEE Computer Society. 1
- [52] Apoorva Rajagopal, Christopher L Dembia, Matthew S Demers, Denny D Delp, Jennifer L Hicks, and Scott L Delp. Full-body musculoskeletal model for muscle-driven simulation of human gait. *IEEE transactions on biomedical engineering*, 63(10):2068–2079, 2016. 1, 3, 5, 7
- [53] David Schneider, Simon Reiß, Marco Kugler, Alexander Jaus, Kunyu Peng, Susanne Sutschet, Muhammad Saquib Sarfraz, Sven Matthiesen, and Rainer Stiefelhagen. Muscles in time: Learning to understand human motion in-depth by simulating muscle activations. *Advances in Neural Information Processing Systems*, 2025. 2, 3, 5
- [54] Robert V Schulte, Marijke Zondag, Jaap H Buurke, and Erik C Prinsen. Multi-day emg-based knee joint torque estimation using hybrid neuromusculoskeletal modelling and convolutional neural networks. *Frontiers in Robotics and AI*, 9:869476, 2022. 3
- [55] Paniz Sedighi, Xingyu Li, and Mahdi Tavakoli. Emg-based intention detection using deep learning for shared control in upper-limb assistive exoskeletons. *IEEE Robotics and Automation Letters*, 2023. 1, 3
- [56] Masashi Sekiya, Sho Sakaino, and Tsuji Toshiaki. Linear logistic regression for estimation of lower limb muscle activations. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 27(3):523–532, 2019. 3
- [57] Antun Skuric, Vincent Padois, Nasser Rezzoug, and David Daneý. On-line feasible wrench polytope evaluation based on human musculoskeletal models: an iterative convex hull method. *IEEE robotics and automation letters*, 7(2):5206–5213, 2022. 1
- [58] Hyungeun Song and Yoichi Hori. Inverse muscle group activity estimation based on neuromusculoskeletal system model. In *TENCON 2015-2015 IEEE Region 10 Conference*, pages 1–5. IEEE, 2015. 3
- [59] Yokhesh K Tamilselvam, Jacky Ganguly, Rajni V Patel, and Mandar Jog. Musculoskeletal model to predict muscle activity during upper limb movement. *Ieee Access*, 9:111472–111485, 2021. 3
- [60] Guy Tevet, Sigal Raab, Brian Gordon, Yonatan Shafir, Daniel Cohen-Or, and Amit Haim Bermano. Human motion diffusion model. In *The Eleventh International Conference on Learning Representations ICLR 2023*, Kigali, Rwanda, 2023. OpenReview.net. 1
- [61] Elly Trepman, Richard E Gellman, Lyle J Micheli, and CARLO J De Luca. Electromyographic analysis of grandplié in ballet and modern dancers. *Medicine and science in sports and exercise*, 30(12):1708–1720, 1998. 3
- [62] J. Wang, Y. Yuan, Z. Luo, K. Xie, D. Lin, U. Iqbal, S. Fidler, and S. Khamis. Learning human dynamics in autonomous driving scenarios. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 20739–20749, Los Alamitos, CA, USA, 2023. IEEE Computer Society. 3
- [63] Keenon Werling, Dalton Omens, Jeongseok Lee, Ioannis Exarchos, and C. Karen Liu. Fast and Feature-Complete Differentiable Physics Engine for Articulated Rigid Bodies with Contact Constraints. In *Robotics: Science and Systems XVII, Virtual Event, July 12-16, 2021*, Virtual, 2021. RSS Foundation. 2, 5, 7
- [64] Keenon Werling, Nicholas A Bianco, Michael Raitor, Jon Stingel, Jennifer L Hicks, Steven H Collins, Scott L Delp, and C Karen Liu. Addbiomechanics: Automating model scaling, inverse kinematics, and inverse dynamics from human motion data through sequential optimization. *Plos one*, 18(11):e0295152, 2023. 2, 5
- [65] Keenon Werling, Janelle Kaneda, Tian Tan, Rishi Agarwal, Six Skov, Tom Van Wouwe, Scott Uhlich, Nicholas Bianco, Carmichael Ong, Antoine Falisse, et al. Addbiomechanics dataset: Capturing the physics of human motion at scale. In *European Conference on Computer Vision*, pages 490–508. Springer, 2024. 2, 3
- [66] Lahiru N Wimalasena, Jonas F Braun, Mohammad Reza Keshtkaran, David Hofmann, Juan Álvaro Gallego, Cristiano Alessandro, Matthew C Tresch, Lee E Miller, and Chethan Pandarinath. Estimating muscle activation from emg using deep learning-based dynamical systems models. *Journal of neural engineering*, 19(3):036013, 2022. 3
- [67] Jungdam Won, Deepak Gopinath, and Jessica Hodgins. Control strategies for physically simulated characters performing two-player competitive sports. *ACM Transactions on Graphics (TOG)*, 40(4):1–11, 2021. 3
- [68] Jungdam Won, Deepak Gopinath, and Jessica Hodgins. Physics-based character controllers using conditional vaes. *ACM Transactions on Graphics (TOG)*, 41(4):1–12, 2022. 3
- [69] Baoping Xiong, Nianyin Zeng, Han Li, Yuan Yang, Yurong Li, Meilan Huang, Wuxiang Shi, Min Du, and Yudong Zhang. Intelligent prediction of human lower extremity joint moment: an artificial neural network approach. *Ieee Access*, 7:29973–29980, 2019. 3
- [70] Aya Yaacoub, Vincent Thomas, Francis Colas, and Pauline Maurice. A probabilistic model for cobot decision making to mitigate human fatigue in repetitive co-manipulation tasks. *IEEE Robotics and Automation Letters*, 2023. 1
- [71] Katsu Yamane, Akihiko Murai, Sadahiro Takaya, and Yoshihiko Nakamura. Muscle tension database for contact-free estimation of human somatosensory information. In *2009 IEEE International Conference on Robotics and Automation*, pages 633–638. IEEE, 2009. 3
- [72] X. Yi, Y. Zhou, M. Habermann, S. Shimada, V. Golyanik, C. Theobalt, and F. Xu. Physical inertial poser (pip): Physics-aware real-time human motion tracking from sparse inertial sensors. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 13157–13168, Los Alamitos, CA, USA, 2022. IEEE Computer Society. 3
- [73] Y. Yuan and K. Kitani. Ego-pose estimation and forecasting as real-time pd control. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages

- 10081–10091, Los Alamitos, CA, USA, 2019. IEEE Computer Society. 3
- [74] Felix E Zajac. Muscle and tendon: properties, models, scaling, and application to biomechanics and motor control. *Critical reviews in biomedical engineering*, 17(4):359–411, 1989. 2
 - [75] Petrisa Zell and Bodo Rosenhahn. A physics-based statistical model for human gait analysis. In *Pattern Recognition*, pages 169–180, Cham, 2015. Springer International Publishing. 3
 - [76] P. Zell and B. Rosenhahn. Learning-based inverse dynamics of human motion. In *2017 IEEE International Conference on Computer Vision Workshop (ICCVW)*, pages 842–850, Los Alamitos, CA, USA, 2017. IEEE Computer Society. 3
 - [77] P. Zell, B. Wandt, and B. Rosenhahn. Joint 3d human motion capture and physical analysis from monocular videos. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 17–26, Los Alamitos, CA, USA, 2017. IEEE Computer Society. 3
 - [78] Petrisa Zell, Bodo Rosenhahn, and Bastian Wandt. Weakly-supervised learning of human dynamics. In *Computer Vision – ECCV 2020*, pages 68–84, Cham, 2020. Springer International Publishing. 3
 - [79] Longbin Zhang, Davit Soselia, Ruoli Wang, and Elena M Gutierrez-Farewik. Estimation of joint torque by emg-driven neuromusculoskeletal models and lstm networks. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2023. 3
 - [80] Yufei Zhang, Jeffrey O Kephart, and Qiang Ji. Incorporating physics principles for precise human motion prediction. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 6164–6174, 2024. 3
 - [81] Chenhui Zuo, Kaibo He, Jing Shao, and Yanan Sui. Self model for embodied intelligence: Modeling full-body human musculoskeletal system and locomotion control with hierarchical low-dimensional representation. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 13062–13069. IEEE, 2024. 3