

Supplementary Materials for “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

A. Licenses

All the data used are from the open-sourced datasets and for research purposes only. We give the links to the gathered datasets here.

- AMASS: <https://amass.is.tue.mpg.de/license.html>
- Muscles in Actions: <https://musclesinaction.cs.columbia.edu/>
- AddBiomechanics: https://addbiomechanics.org/download_data.html
- Muscles in Time: <https://davidschneider.ai/mint/>
- ImDy: <https://foruck.github.io/ImDy/>

The subfigures of “Activation Dynamics” and “Contraction Dynamics” in Figure 1 are borrowed from Uchida, Thomas K., and Scott L. Delp. Biomechanics of movement: the science of sports, robotics, and rehabilitation. MIT Press, 2021. Figure 4.16 and Chapter 5.

B. Extensive Experiments

B.1. Analysis on Parameters

We compare the size of the models involved in Table 1 in Table 1. The full HDyS is comparable in #param compared with previous efforts. In addition, it could process four heterogeneous kinematics representations and four heterogeneous dynamics representations, which could not be fulfilled with previous efforts. Moreover, even with a much smaller model scale, HDyS-32D and HDyS-64D manage to provide competitive performances, validating the efficacy of heterogeneous knowledge.

B.2. Extensive Results on Inverse Dynamics

B.2.1 Data Construction

To decompose the contributions of scale and heterogeneity, we construct two sets of control experiments. The first set

Table 1. Model size comparison.

Models	#params
MiA	5.4M
ImDyS	4.0M
HDyS	3.9M
HDyS-32D	0.6M
HDyS-64D	1.4M

of control experiments were controlled for the same data scale, and they differed only in whether the data constituted heterogeneity or not. The second set of control experiments varies only in the scale of the data.

Thus, we constructed HDyS-50/50 to form the first set of control experiments with the original HDyS-Single, and HDyS-Single-50 to form the second set of control experiments with HDyS-Single. In this way, HDyS-Single-50 and HDyS-50/50 formed a third control experiment with the same data from the target dataset, in which homogeneous knowledge in heterogeneous data can be observed. To construct the other datasets part of HDyS-50/50, we proportionally sampled the training data from other datasets so that the total amount of data selected was equal to 50% of the total amount of the target dataset. The details of the construction are shown in Tab. 2.

B.2.2 More Ablation Studies

An additional ablation study is provided to evaluate the transformer-based temporal refinement. We remove the temporal transformer in the ID decoder and report its performance in Tab. 3. As shown, substantial performance degradation is observed, validating the refinement of the temporal transformer.

*Corresponding authors.

Table 2. Composition of training data in Table 2.

Target dataset	Model	#seq for training				
		AddBiomechanics	MiA	ImDy	MinT	AMASS
AddBiomechanics	HDyS-Single-50	5810	-	-	-	-
	HDyS-50/50	5810	601	3381	111	1212
	HDyS-Single	11621	-	-	-	-
MiA	HDyS-Single-50	-	2446	-	-	-
	HDyS-50/50	526	2446	1246	41	632
	HDyS-Single	-	4891	-	-	-

Table 3. Ablation study on the transformer-based temporal refinement.

Methods	ImDy	AddBiomechanics	MinT		MiA	
	mPJE↓	mPJE↓	RMSE↓	PCC↑	RMSE↓	PCC↑
	avg/bst	avg/bst	avg/bst	avg/bst	avg/bst	avg/bst
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
HDyS w/o Temporal Refinement	0.7002/0.5334	0.1393/0.1489	0.0666/0.0670	0.7372/0.7325	15.4/15.1	0.5748/0.5788

Table 4. More ablative baselines on GroundLink.

Methods	HDyS-Marker	HDyS-SMPL	HDyS-keypoint	HDyS
L-Foot <i>mPJE</i> ↓	0.0673	0.0591	0.0584	0.0514
R-Foot <i>mPJE</i> ↓	0.0930	0.0732	0.1047	0.0694

B.2.3 Architectural Clarification and Justification

Our basic idea is to use basic structures wherever possible to highlight the power of inherent homogeneity. Therefore, we tend to use basic three-layer MLPs for single-frame fixed-size inputs (like joint angles) while maintaining non-linearity modeling ability. Transformers are adopted when variable-size inputs (like markers and joints) or sequential inputs (in the ID decoder) are used. The numbers of hidden dimensions and attention heads are designed to match the dimensions of inputs/outputs. The number of transformer layers is selected to match the number of parameters of existing baselines as listed in Appendix B.1. While we believe HDyS could be enhanced by more sophisticated architectures like an auto-regressive operation manner, we leave this for future work.

B.3. More Analysis on Ground Reaction Force Prediction

In Tab. 4, we include some ablative baselines for the influence of different kinematics representations on GRF estimation, validating the mutual benefit of unifying kinematics representations again.

B.4. More Analysis on Biomechanical Human Simulation

Quantitative results are shown in Tab. 5. As shown, increasing the simulation frame rate effectively reduces the simulation error. And HDyS consistently provides competitive

Table 5. Extended results reported in per-frame MSE on biomechanical human simulation.

Methods	90FPS	120FPS	150FPS
HDyS-2-steps	0.1860	0.0591	0.0244
Optimized-2-steps	0.1909	0.0607	0.0253
HDyS-3-steps	1.7118	0.5257	0.2125
Optimized-3-steps	1.8306	0.5495	0.2223
HDyS-4-steps	1.8651	1.5721	1.1173
Optimized-4-steps	2.0106	1.7630	0.7081
HDyS-5-steps	2.2233	2.1384	2.0482
Optimized-5-steps	2.6027	2.5147	2.5017

Table 6. Hyperparameters for two primitives. σ : fixed variance for policy. γ : discount factor. ϵ : clip range for PPO

	σ	γ	ϵ
Value	0.05	0.99	0.2

performances. However, drifting errors could still be observed.

B.5. Details of Physical Character Control

We exclude all motion sequences involving sitting on chairs, walking on treadmills, leaning on tables, stepping on stairs, or floating in the air. This filtering process yields a dataset comprising 10,047 high-quality motion sequences for training and 140 sequences for testing. Following the PHC setting, as a baseline comparison, we trained two single primitives to demonstrate that HDyS enhances physical character control performance. Each primitive is implemented as a six-layer MLP with units [2048, 1536, 1024, 1024, 512, 512] and employs SiLU as the activation function. HDyS latents corresponding to key points are incor-

porated as additional observations. The only difference between the two primitives lies in the input, one without HDyS latents denoted as *Baseline*, and the other one with HDyS latents denoted as *Baseline w/ HDyS*. For training, we employ the Adam optimizer with a learning rate of $2e-5$, a batch size of 768, and train the model for 10,000 steps. The hyperparameters used during training can be found in Table 6.

References