3D Human Pose Estimation with Muscles

Kevin Zhu AliAsghar MohammadiNasrabadi Alexander Wong John McPhee

UNIVERSITY OF WATERLOO | FACULTY OF ENGINEERING

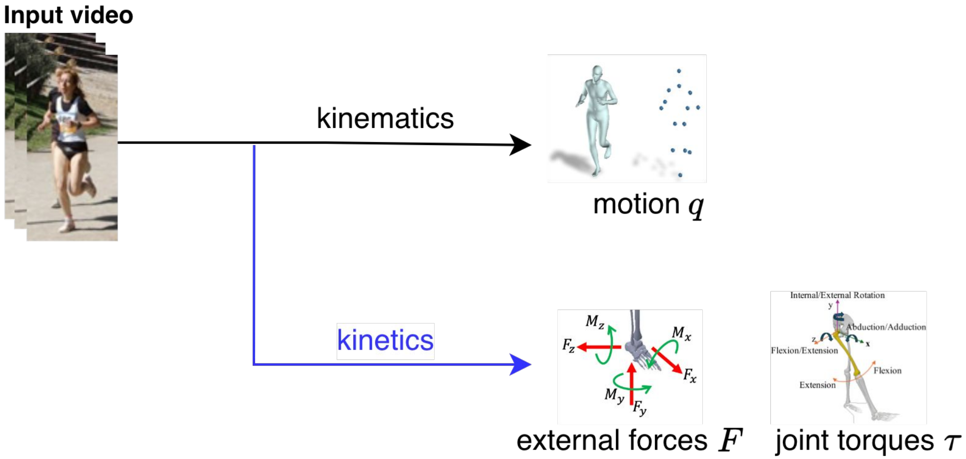
MOTION RESEARCH GROUP

Monocular physics-based human pose estimation

Physics-based human pose estimation (PHPE) aims to reduce motion artifacts and improve contact estimation by enforcing the underlying dynamics equations:

$$M(q, A)\ddot{q} + C\big(q, \dot{q}\big) = \tau + F$$

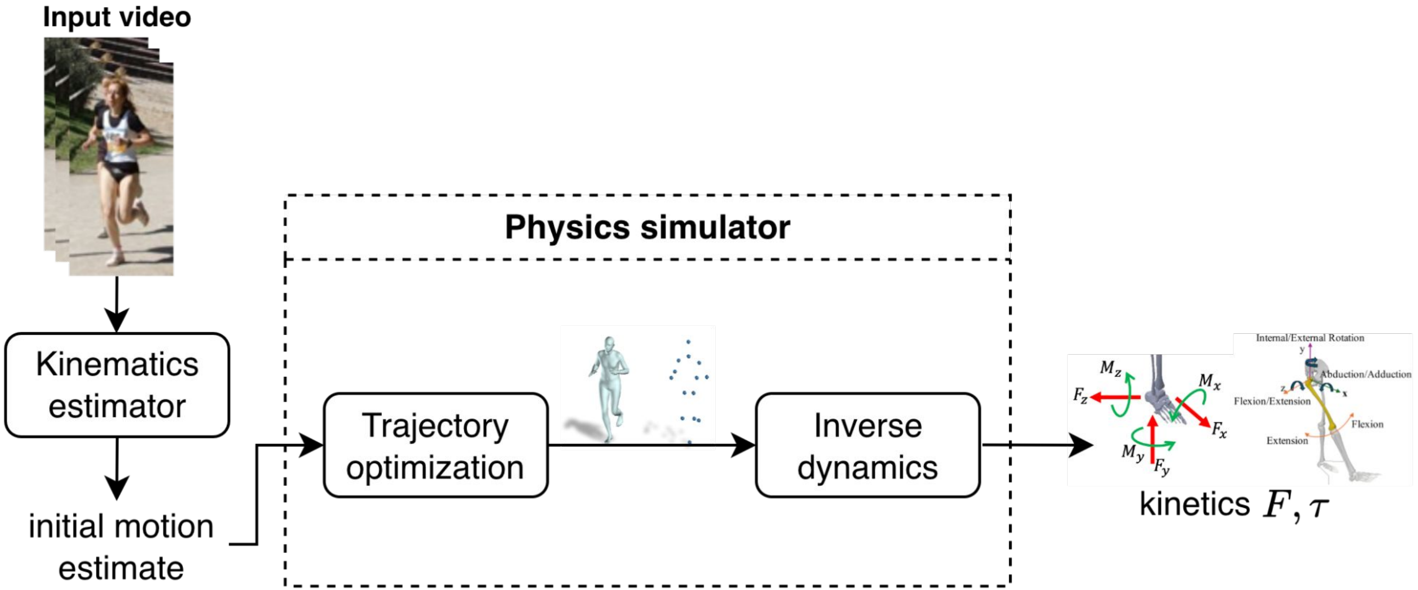
Hence, in addition to human kinematics (motion), we need to estimate the corresponding kinetics (external forces, joint torques, etc).



Optimization-based PHPE [20,19,77,64,41,82,59]

<u>Method 1 (with a physics simulator):</u>

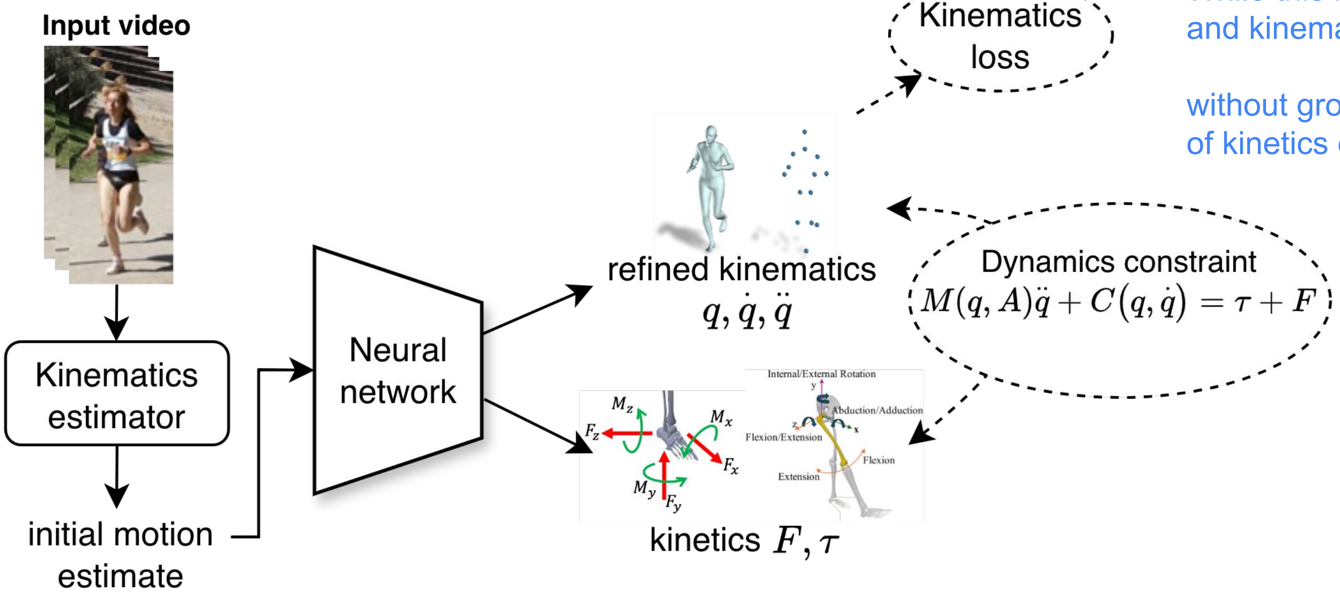
Reconstruct the predicted motion using trajectory optimization with an external physics simulator.



Regression-based PHPE [85,37,63]

<u>Method 2 (without a physics simulator):</u>

Directly regress kinetics, while enforcing a dynamics loss by minimizing residual force.

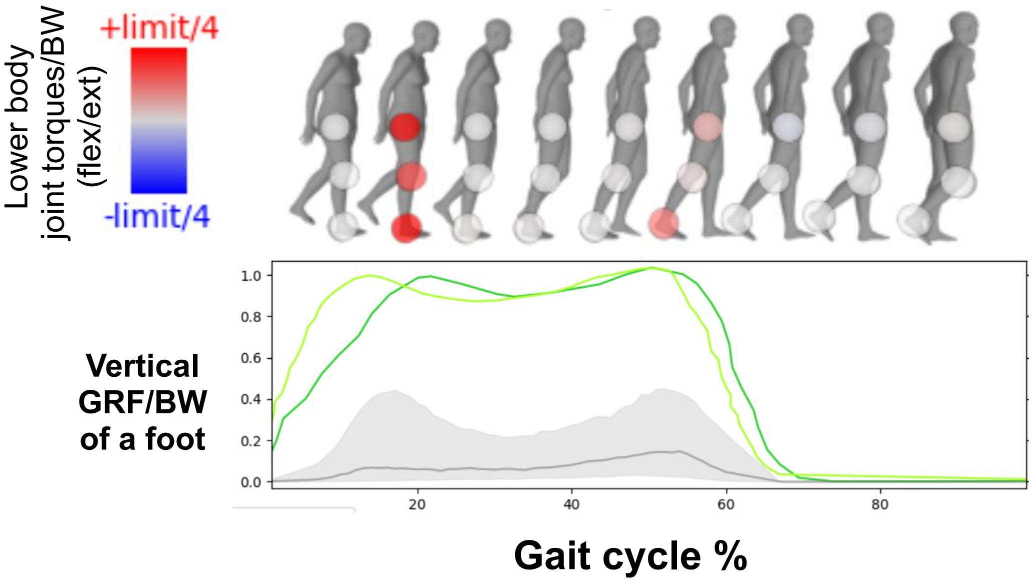


While this improves contact estimation and kinematic plausibility metrics,

without ground-truth force labels, reliability of kinetics estimation unknown.

Examples of implausible kinetics (gait)

Regressed kinetics from a **SOTA pose estimator** [85] vs **biomechanics** references [72,18,76,8,29,75,55,80,62],

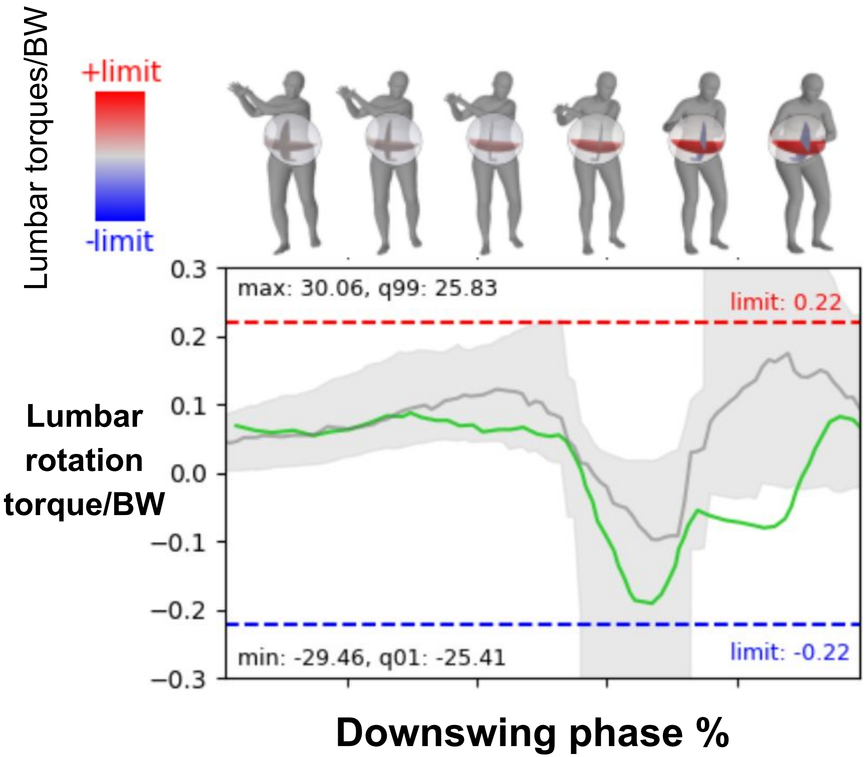


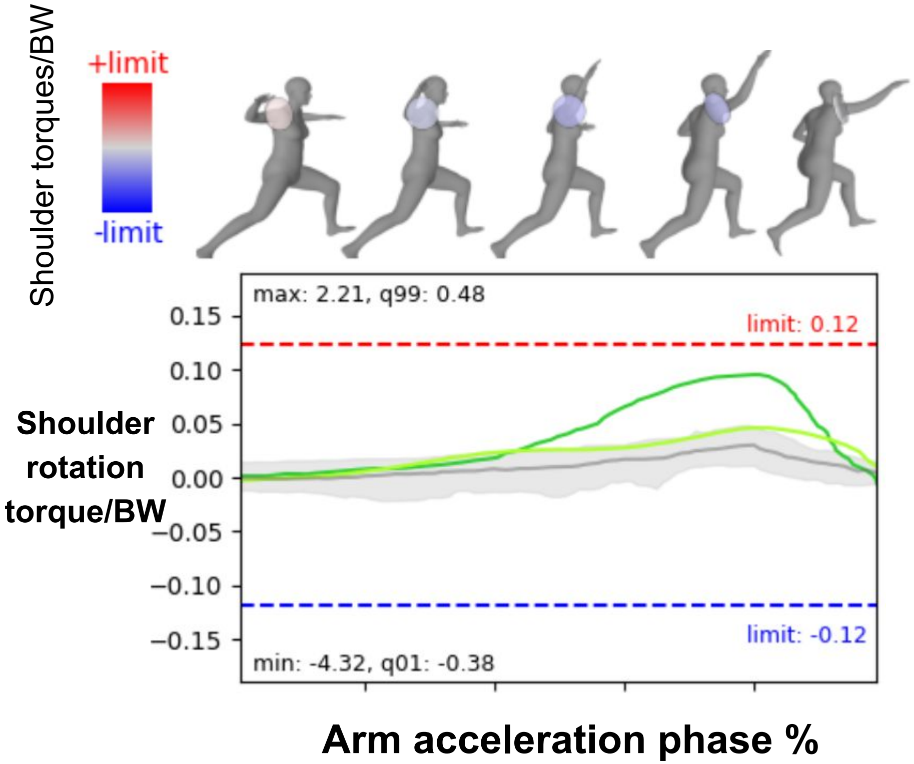
Gray: regressed kinetics from PHPE [85] (median with 25-75% quantile band).

All values scaled by body weight (BW)

Green curves are values collected from other biomechanics studies for comparison of trends and magnitudes.

Examples of implausible kinetics (golf, baseball)

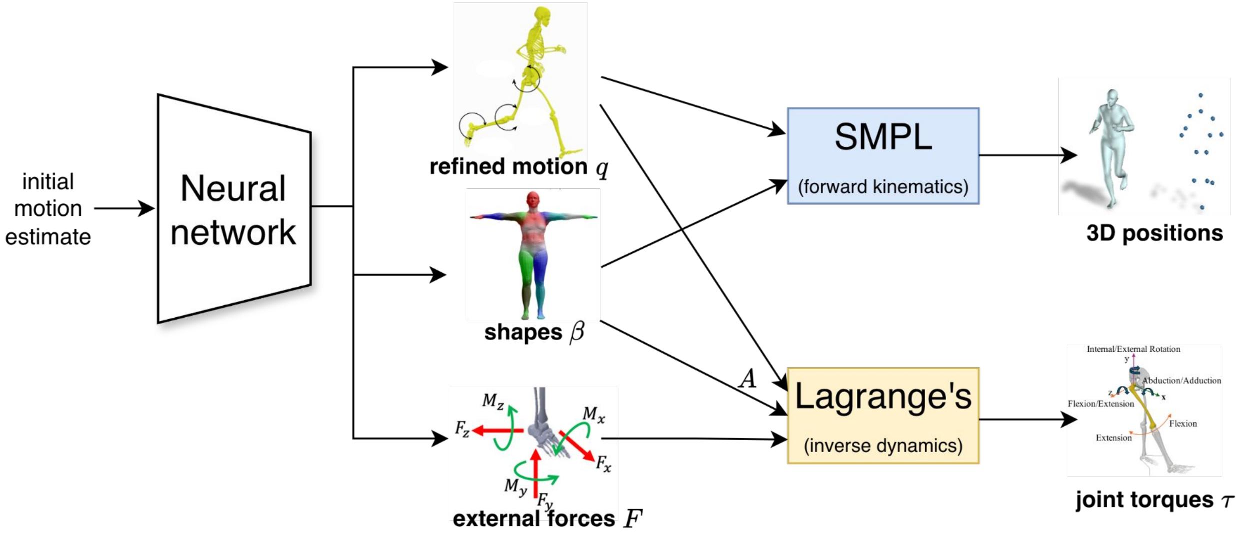




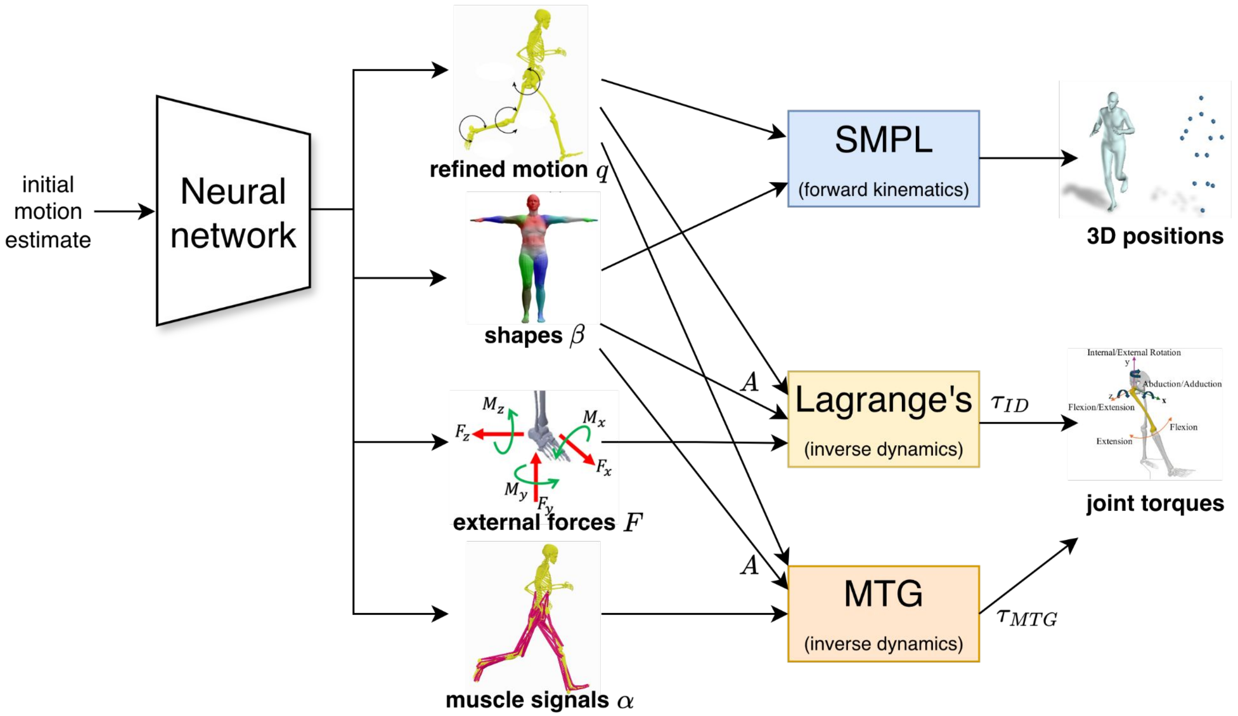
Regression-based PHPE

Problem
Kinetics prediction (RHS) is underconstrained.

$$M(q, A)\ddot{q} + C(q, \dot{q}) = \tau + F$$



MusclePose



Problem

Kinetics prediction (RHS) is underconstrained.

$$M(q, A)\ddot{q} + C(q, \dot{q}) = \tau + F$$

Proposed

Add a parametric MTG module as another way to predict torques.

$$L_\tau = |\tau_{ID} - \tau_{MTG}|$$

(i.e. use MTGs as a regularizer)

Muscle torque generators (MTGs) [51,25]

Motivation:

- **simplicity**: MTGs are simplified muscle models that approximate torque at the joint level from joint kinematics (q), muscle activation signals ($\boldsymbol{\alpha}$), and anthropometric features (A).

joint kinematics q, activations α, anthropometrics A $\longrightarrow$ $\boxed{\text{MTG}}$ $\longrightarrow$ τ_{MTG}

- **compatibility**: MTGs consist of parametric equations that are differentiable and can be directly incorporated into existing learning frameworks.

muscle activation $\alpha \in [0, 1]$

peak isokinetic torque $\tau_0(A)$

$$\tau_{MTG} = \tau_{active} + \tau_{passive}$$

$$\tau_{active} = \alpha \tau_\omega \tau_\theta \tau_0$$

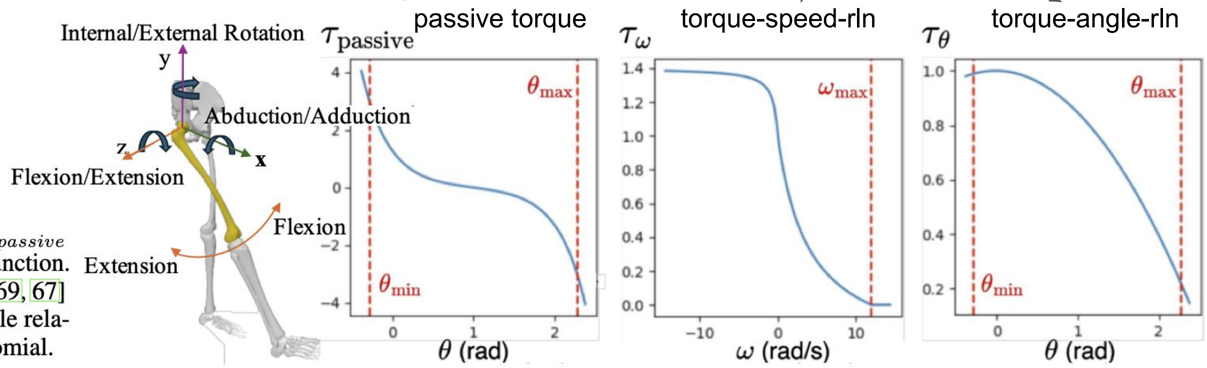
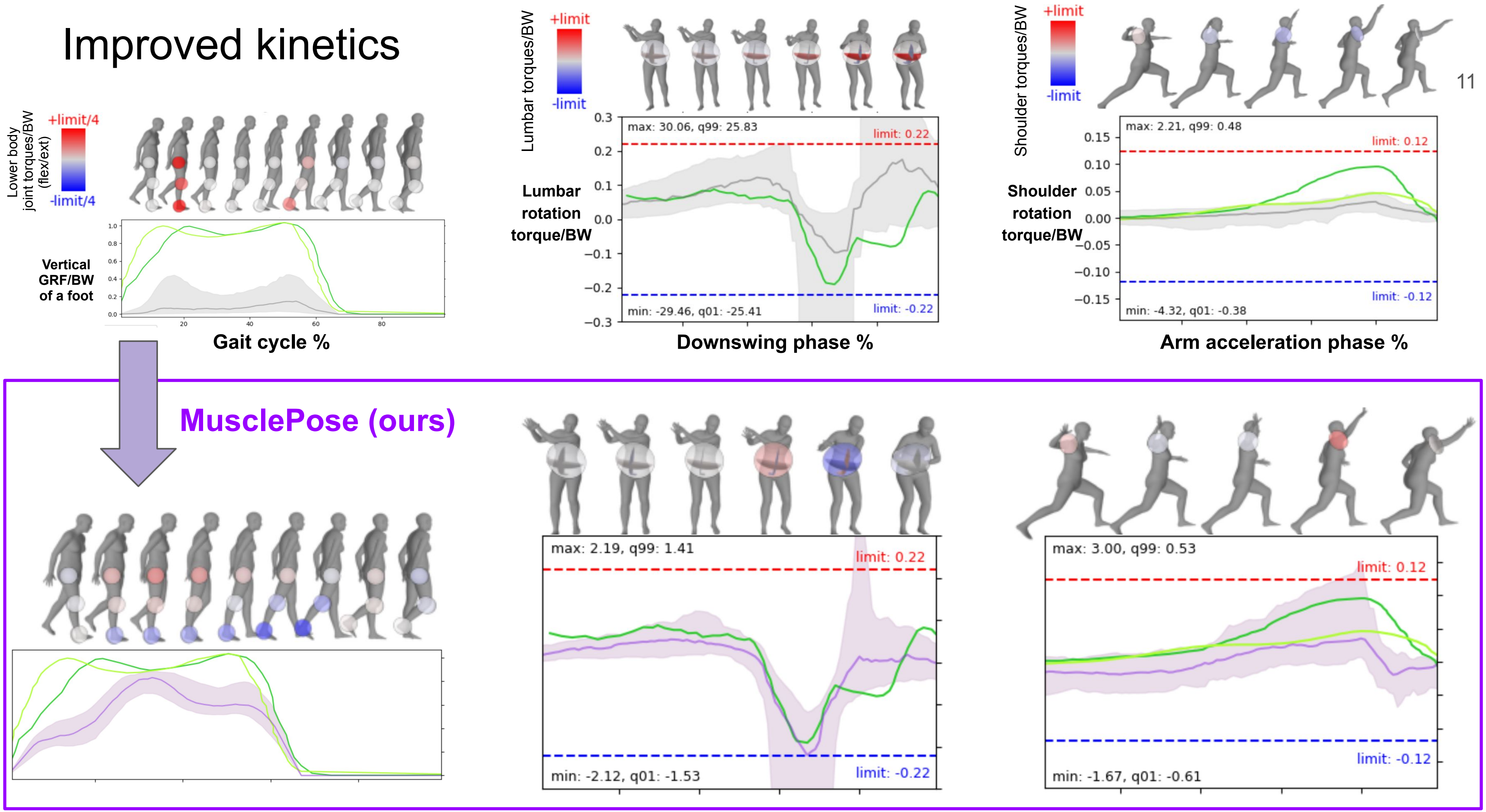


Figure 2: Examples of MTG curves for hip flexion. $\tau_{passive}$ models the passive torque [79] as a double exponential function. τ_ω models the active-torque–angular-speed relationship [69, 67] as a piecewise function. τ_θ models the active-torque-angle relationship [21, 33] as the non-negative portion of a polynomial.

Results

Improved kinetics



Kinetics evaluation

		Kinetic plausibility		$\mathrm{GRF}_v^{/W}$			$\mathrm{GRF}_v^{/W}$ (%f)		
		$\mathbf{F}_{res}^{/W}\downarrow$	MTV	{med	q99	max}	< {0.01,	0.1,	0.5}
†H36M	CLIFF	-	-	-			-		
	PhysPT	0.4	5.3	{0.4	2.4	10.0}	{7.2	20.8	60.0}
	MusclePose	**0.1**	2.5	{1.0	1.2	3.0}	{3.0	3.0	5.2}
†3DPWoc	CLIFF	-	-		-			-	
	PhysPT	0.9	27.0	{0.5	1.2	3.9}	{5.3	11.8	52.4}
	MusclePose	**0.3**	12.1	{1.0	1.6	4.3}	{5.3	5.3	6.3}

Fres: mean residual force scaled by body weight.

MTV (mean torque variation): mean absolute change in joint torques over consecutive frames (in Newton*metres/frame) to assess torque continuity.

GRFv %f: percent of frames where the total vertical GRF is below (1, 10, 50% body weight).

Kinematics evaluation

		Pos.	Kinematic plausibility			Float (%f)		
		MJE↓	ACC↓	FS	GP	$\mathcal{H}_{min} > \{1, 10, 20\}$mm		
†H36M	CLIFF	**46.5**	26.3	-	-		-	
	PhysPT	50.6	13.7	34.7	6.8	{59.0	31.6	8.5}
	MusclePose	48.4	**12.9**	37.2	26.0	{8.0	3.0	1.3}
†3DPWoc	CLIFF	**24.0**^{P}	13.8^P	-	-		-	
	PhysPT	25.9^P	**3.0**^{P}	7.8	11.2	{82.9	73.9	57.2}
	MusclePose	27.6^P	4.3^P	12.8	30.8	{6.0	4.7	3.7}

MJE: mean joint positional error in mm.

ACC (acceleration loss): mean L2 norm in mm/frame2 between the predicted and ground truth keypoint accelerations to access jitter.

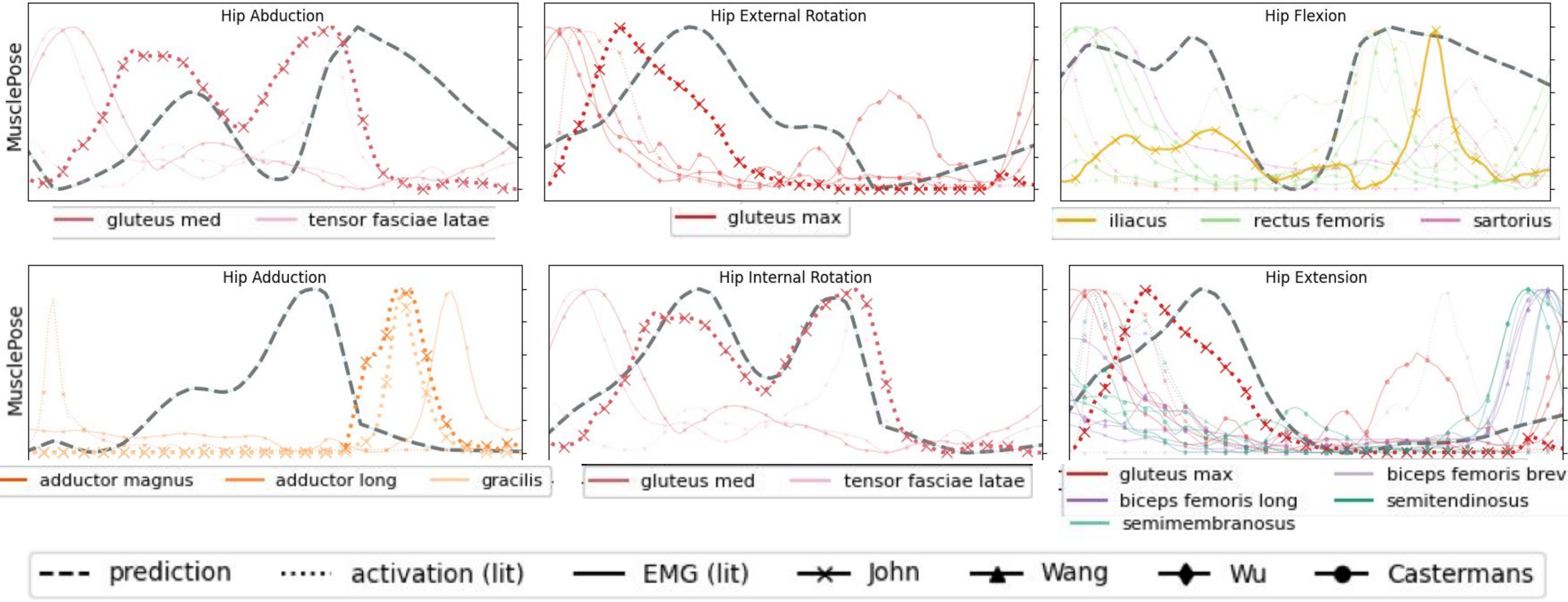
FS (foot skating): average displacement in mm of vertices in contact with the ground in consecutive frames.

GP (ground penetration) is the average vertical distance to the ground in mm of vertices below the ground.

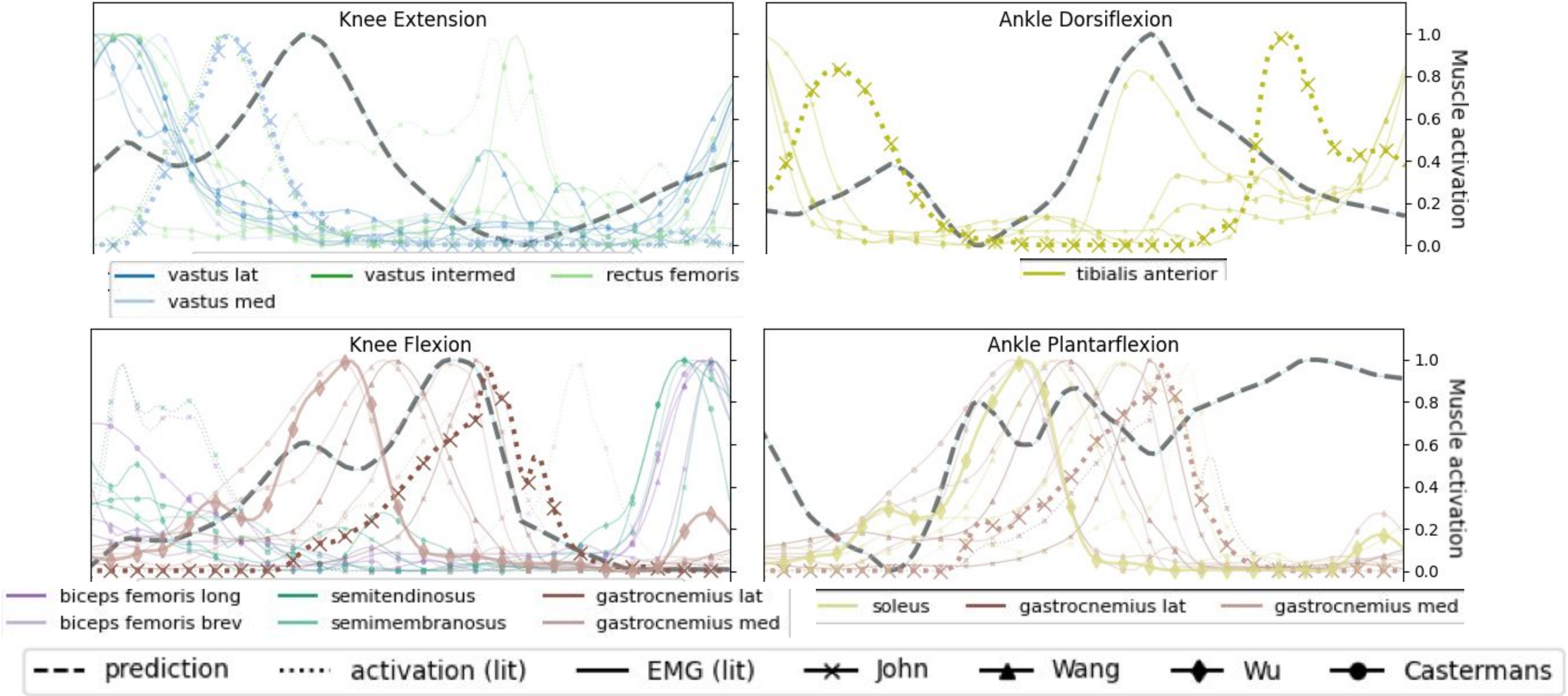
Floating %f: percent of frames the minimum vertex height is above (1, 10, 20mm).

Inferred muscle signals (gait – hip)

Mean muscle activation signal predictions (dashed) compared to activations (dotted) and EMG data (solid) collected from other gait studies [72,76,8,29]. Min-max scaling applied to all.



Inferred muscle signals (gait – knee, ankle)



Summary of contributions

We introduced MusclePose as a regression-based PHPE to infer human dynamics from monocular videos,

- this is the first pose estimator to incorporate muscle dynamics modeling, predict muscle activation signals, and predict detailed human anthropometrics,
- we showed improvements in biofidelity, while maintaining competitive positional accuracy.

Thanks!

References

[1] D. E. Anderson, M. L. Madigan, and M. A. Nussbaum. Maximum voluntary joint torque as a function of joint angle and angular velocity: Model development and application to the lower limb. *Journal of Biomechanics*, 40(14):3105–3113, 2007.

[2] Y. Bengio, J. Louradour, R. Collobert, and J. Weston. Curriculum learning. In *Proceedings of the 26th Annual International Conference on Machine Learning*, ICML '09, page 41–48, 2009.

[3] N. A. Bernshtein. The co-ordination and regulation of movements. 1967.

[4] C. Brown, W. McNally, and J. McPhee. Optimal control of joint torques using direct collocation to maximize ball carry distance in a golf swing. *Multibody System Dynamics*, 50(3), 2020.

[5] C. Brown and J. J. McPhee. Predictive forward dynamic simulation of manual wheelchair propulsion on a rolling dynamometer. *Journal of biomechanical engineering*, 2020.

[6] A. Buchmann, S. Wenzler, L. Welte, and D. Renjewski. The effect of including a mobile arch, toe joint, and joint coupling on predictive neuromuscular simulations of human walking. *Scientific Reports*, 2024.

[7] Y. Cai, L. Ge, J. Liu, J. Cai, T. J. Cham, J. Yuan, and N. M. Thalmann. Exploiting spatial-temporal relationships for 3D pose estimation via graph convolutional networks. In *Proceedings of the IEEE International Conference on Computer Vision*, volume 2019-October, 2019.

[8] T. Castermans, M. Duvinage, G. Cheron, and T. Dutoit. Towards effective non-invasive brain-computer interfaces dedicated to gait rehabilitation systems. *Brain Sciences*, 4(1):1–48, 2014.

[9] S.-W. Chen, W.-T. Tang, J.-T. Kung, T.-Y. Hung, W.-H. Lin, Y.-L. Chen, and D. J. Burgee. Comparison of ground reaction force among stride types in baseball pitching. *Sports Biomechanics*, 0(0):1–14, 2024.

[10] Y. Cheng, B. Yang, B. Wang, and R. T. Tan. 3D human pose estimation using spatio-temporal networks with explicit occlusion training. In *AAAI 2020 - 34th AAAI Conference on Artificial Intelligence*, 2020.

[11] H. Ci, C. Wang, X. Ma, and Y. Wang. Optimizing network structure for 3D human pose estimation. In *Proceedings of the IEEE International Conference on Computer Vision*, volume 2019-October, 2019.

[12] D. J. Cleather and A. M. J. Bull. Lower-extremity musculoskeletal geometry affects the calculation of patellofemoral forces in vertical jumping and weightlifting. *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, 224:1073 – 1083, 2010.

[13] B. Danaei and J. J. McPhee. Model-based acetabular cup orientation optimization based on minimizing the risk of edge-loading and implant impingement following total hip arthroplasty. *Journal of biomechanical engineering*, 2022.

[14] R. Drillis, R. Contini, and M. Bluestein. Body segment parameters. *Artificial limbs*, 8(1):44–66, 1964.

[15] G. A. Dudley, R. T. Harris, M. R. Duvoisin, B. M. Hather, and P. Buchanan. Effect of voluntary vs. artificial activation on the relationship of muscle torque to speed. *Journal of Applied Physiology*, 69(6), 1990.

[16] R. Dumas, L. Chèze, and J. P. Verriest. Adjustments to mcconville et al. and young et al. body segment inertial parameters. *Journal of biomechanics*, 40 3:543–53, 2007.

[17] R. Featherstone. *Rigid Body Dynamics Algorithms*. 2008.

[18] C. A. Fukuchi, R. K. Fukuchi, and M. Duarte. A public dataset of overground and treadmill walking kinematics and kinetics in healthy individuals. *PeerJ*, 6, 2018.

[19] E. Gartner, M. Andriluka, H. Xu, and C. Sminchisescu. Trajectory Optimization for Physics-Based Reconstruction of 3d Human Pose from Monocular Video. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2022-June, 2022.

[20] E. Gärtner, M. Andriluka, E. Coumans, and C. Sminchisescu. Differentiable dynamics for articulated 3d human motion reconstruction. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2022.

[21] D. Haering, C. Pontonnier, N. Bideau, G. Nicolas, and G. Dumont. Using Torque-Angle and Torque-Velocity Models to Characterize Elbow Mechanical Function: Modeling and Applied Aspects. *Journal of Biomechanical Engineering*, 141(8), 2019.

[22] N. Haraguchi, A. Nasr, K. A. Inkol, K. Hase, and J. McPhee. Human and passive lower-limb exoskeleton interaction analysis: Computational study with dynamics simulation using nonlinear model predictive control. *2023 62nd Annual Conference of the Society of Instrument and Control Engineers (SICE)*, pages 844–849, 2023.

[23] A. V. Hill. The heat of shortening and the dynamic constants of muscle. *Proceedings of The Royal Society B: Biological Sciences*, 126:136–195, 1938.

[24] P. D. Hoang, R. B. Gorman, G. Todd, S. C. Gandevia, and R. D. Herbert. A new method for measuring passive length-tension properties of human gastrocnemius muscle in vivo. *Journal of Biomechanics*, 38(6):1333–1341, 2005.

[25] K. A. Inkol, C. Brown, W. McNally, C. Jansen, and J. McPhee. Muscle torque generators in multibody dynamic simulations of optimal sports performance. *Multibody System Dynamics*, 50(4), 2020.

[26] K. A. Inkol and J. J. McPhee. Using dynamic simulations to estimate the feasible stability region of feet-in-place balance recovery for lower-limb exoskeleton users. *2022 9th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob)*, pages 1–6, 2022.

[27] C. Ionescu, D. Papava, V. Olaru, and C. Sminchisescu. Human3.6M: Large scale datasets and predictive methods for 3D human sensing in natural environments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(7), 2014.

[28] C. Jansen and J. J. McPhee. Predictive dynamic simulation of olympic track cycling standing start using direct collocation optimal control. *Multibody System Dynamics*, 49:53–70, 2020.

[29] C. T. John, F. C. Anderson, J. S. Higginson, and S. L. D. and. Stabilisation of walking by intrinsic muscle properties revealed in a three-dimensional muscle-driven simulation. *Computer Methods in Biomechanics and Biomedical Engineering*, 16(4):451–462, 2013.

[30] S. Johnson and M. Everingham. Clustered pose and nonlinear appearance models for human pose estimation. In *British Machine Vision Conference*, 2010.

[31] A. Kanazawa, M. J. Black, D. W. Jacobs, and J. Malik. End-to-End Recovery of Human Shape and Pose. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2018.

[32] B. Katz. The relation between force and speed in muscular contraction. *The Journal of Physiology*, 96, 1939.

[33] M. A. King, C. Wilson, and M. R. Yeadon. Evaluation of a torque-driven model of jumping for height. *Journal of Applied Biomechanics*, 22(4), 2006.

[34] M. Kocabas, N. Athanasiou, and M. J. Black. Vibe: Video inference for human body pose and shape estimation. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2020.

[35] N. Kolotouros, G. Pavlakos, M. J. Black, and K. Daniilidis. Learning to reconstruct 3d human pose and shape via model-fitting in the loop. In *Proceedings of the International Conference on Computer Vision*, 2019.

[36] N. Kolotouros, G. Pavlakos, and K. Daniilidis. Convolutional mesh regression for single-image human shape reconstruction. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4496–4505, 2019.

[37] J. Li, S. Bian, C. Xu, G. Liu, G. Yu, and C. Lu. D &D: Learning Human Dynamics from Dynamic Camera. In *ECCV*, 2022.

[38] J. Li, C. Xu, Z. Chen, S. Bian, L. Yang, and C. Lu. Hybrik: A hybrid analytical-neural inverse kinematics solution for 3d human pose and shape estimation. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2021.

[39] W. Li, H. Liu, H. Tang, P. Wang, and L. Van Gool. MHFormer: Multi-Hypothesis Transformer for 3D Human Pose Estimation. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2022-June, 2022.

[40] Z. Li, J. Liu, Z. Zhang, S. Xu, and Y. Yan. Cliff: Carrying location information in full frames into human pose and shape estimation. In *European Conference on Computer Vision*, 2022.

[41] Z. Li, J. Sedlar, J. Carpentier, I. Laptev, N. Mansard, and J. Sivic. Estimating 3D motion and forces of person-object interactions from monocular video. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2019-June, 2019.

[42] K. Lin, L. Wang, and Z. Liu. Mesh graphormer. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 12919–12928, 2021.

[43] D. M. Lindsay and J. F. Horton. Trunk rotation strength and endurance in healthy normals and elite male golfers with and without low back pain. *North American journal of sports physical therapy : NAJSPT*, 1 2:80–9, 2006.

[44] I. Loshchilov and F. Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2017.

[45] T. Luan, Y. Wang, J. Zhang, Z. Wang, Z. Zhou, and Y. Qiao. Pc-hmr: Pose calibration for 3d human mesh recovery from 2d images/videos. In *AAAI Conference on Artificial Intelligence*, 2021.

[46] N. Mahmood, N. Ghorbani, N. F. Troje, G. Pons-Moll, and M. Black. AMASS: Archive of motion capture as surface shapes. In *Proceedings of the IEEE International Conference on Computer Vision*, volume 2019-October, 2019.

[47] J. Martinez, R. Hossain, J. Romero, and J. J. Little. A Simple Yet Effective Baseline for 3d Human Pose Estimation. In *Proceedings of the IEEE International Conference on Computer Vision*, volume 2017-October, 2017.

[48] W. McNally and J. McPhee. Dynamic Optimization of the Golf Swing Using a Six Degree-of-Freedom Biomechanical Model. 2018.

[49] W. J. McNally and J. J. McPhee. Dynamic optimization of the golf swing using a six degree-of-freedom biomechanical model. 2018.

[50] D. Mehta, H. Rhodin, D. Casas, P. Fua, O. Sotnychenko, W. Xu, and C. Theobalt. Monocular 3D human pose estimation in the wild using improved CNN supervision. In *Proceedings - 2017 International Conference on 3D Vision, 3DV 2017*, 2018.

[51] M. Millard, T. K. Uchida, A. Seth, and S. L. Delp. Flexing computational muscle: modeling and simulation of musculotendon dynamics. *Journal of biomechanical engineering*, 135 2:021005, 2013.

[52] G. Moon and K. M. Lee. I2l-meshnet: Image-to-lixel prediction network for accurate 3d human pose and mesh estimation from a single rgb image. *ArXiv*, abs/2008.03713, 2020.

[53] A. Nasr, A. Hashemi, and J. McPhee. Scalable musculoskeletal model for dynamic simulations of upper body movement. *Computer Methods in Biomechanics and Biomedical Engineering*, 2023.

[54] A. Nasr and J. McPhee. Scalable musculoskeletal model for dynamic simulations of lower body movement. *Computer methods in biomechanics and biomedical engineering*, pages 1–27, 2024.

[55] S. Nesbit. Development of a full-body biomechanical model of the golf swing. *International Journal of Modelling and Simulation*, 27(4):392–404, 2007.

[56] A. Newell, P. Hu, L. Lipson, S. R. Richter, and V. Koltun. Comotion: Concurrent multi-person 3d motion. In *ICLR*, 2025.

[57] G. Pavlakos, X. Zhou, K. G. Derpanis, and K. Daniilidis. Coarse-to-fine volumetric prediction for single-image 3D human pose. In *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, volume 2017-January, 2017.

[58] D. Pavllo, C. Feichtenhofer, D. Grangier, and M. Auli. 3D human pose estimation in video with temporal convolutions and semi-supervised training. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2019-June, 2019.

[59] X. B. Peng, P. Abbeel, S. Levine, and M. Van De Panne. DeepMimic: Example-guided deep reinforcement learning of physics-based character skills. *ACM Transactions on Graphics*, 37(4), 2018.

[60] E. D. Pieri, M. Lund, A. Gopalakrishnan, K. P. Rasmussen, D. E. Lunn, and S. J. Ferguson. Refining muscle geometry and wrapping in the tlem 2 model for improved hip contact force prediction. *PLoS ONE*, 13, 2018.

References

[61] D. Rempe, L. J. Guibas, A. Hertzmann, B. Russell, R. Villegas, and J. Yang. Contact and Human Dynamics from Monocular Video. In *19th ACM SIGGRAPH / Eurographics Symposium on Computer Animation 2020, SCA 2020 - Showcases*, 2020.

[62] M. Sabick, M. Torry, Y.-K. Kim, and R. Hawkins. Humeral torque in professional baseball pitchers. *The American journal of sports medicine*, 32:892–8, 07 2004.

[63] S. Shimada, V. Golyanik, W. Xu, P. Pérez, and C. Theobalt. Neural monocular 3D human motion capture with physical awareness. *ACM Transactions on Graphics*, 40(4), 2021.

[64] S. Shimada, V. Golyanik, W. Xu, and C. Theobalt. Phys Cap. *ACM Transactions on Graphics*, 39(6), 2020.

[65] S. Shin, J. Kim, E. Halilaj, and M. J. Black. Wham: Reconstructing world-grounded humans with accurate 3d motion. In *CVPR*, 2024.

[66] T. Siebert, C. Rode, W. Herzog, O. Till, and R. Blickhan. Nonlinearities make a difference: comparison of two common hill-type models with real muscle. *Biological Cybernetics*, 98:133–143, 2008.

[67] E. J. Sprigings. Simulation of the force enhancement phenomenon in muscle. *Computers in Biology and Medicine*, 16(6), 1986.

[68] S. Tripathi, L. Müller, C.-H. P. Huang, O. Taheri, M. J. Black, and D. Tzionas. 3d human pose estimation via intuitive physics. *ArXiv*, abs/2303.18246, 2023.

[69] A. J. van Soest and M. F. Bobbert. The contribution of muscle properties in the control of explosive movements. *Biological Cybernetics*, 69(3), 1993.

[70] G. Varol, D. Ceylan, B. C. Russell, J. Yang, E. Yumer, I. Laptev, and C. Schmid. Bodynet: Volumetric inference of 3d human body shapes. *ArXiv*, abs/1804.04875, 2018.

[71] T. von Marcard, R. Henschel, M. Black, B. Rosenhahn, and G. Pons-Moll. Recovering accurate 3d human pose in the wild using imus and a moving camera. In *European Conference on Computer Vision (ECCV)*, sep 2018.

[72] H. Wang, A. Basu, G. Durandau, and M. Sartori. Comprehensive Kinetic and EMG Dataset of Daily Locomotion with 6 types of Sensors, May 2022.

[73] H. Wang, A. Basu, G. Durandau, and M. Sartori. A wearable real-time kinetic measurement sensor setup for human locomotion. *Wearable Technologies*, Feb. 2023.

[74] J. Wang, S. Yan, Y. Xiong, and D. Lin. Motion Guided 3D Pose Estimation from Videos. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 12358 LNCS, 2020.

[75] D. A. Winter. *Biomechanics and motor control of human movement*. John Wiley & Sons, Hoboken, NJ, USA, 4th edition, 2009.

[76] A. R. Wu, F. Dzeladini, T. J. H. Brug, F. Tamburella, N. L. Tagliamonte, E. H. F. van Asseldonk, H. van der Kooij, and A. J. Ijspeert. An adaptive neuromuscular controller for assistive lower-limb exoskeletons: A preliminary study on subjects with spinal cord injury. *Frontiers in Neurorobotics*, Volume 11 - 2017, 2017.

[77] K. Xie, T. Wang, U. Iqbal, Y. Guo, S. Fidler, and F. Shkurti. Physics-based Human Motion Estimation and Synthesis from Videos. In *Proceedings of the IEEE International Conference on Computer Vision*, 2021.

[78] G. T. Yamaguchi. *Dynamic Modeling of Musculoskeletal Motion*. 2001.

[79] G. T. Yamaguchi. *Dynamic modeling of musculoskeletal motion: A vectorized approach for biomechanical analysis in three dimensions*. Springer, Boston, MA, USA, 1 edition, 2006.

[80] K. Yoichi, E. Sato, and T. Yamaji. Biomechanical analysis of the pitching characteristics of adult amateur baseball pitchers throwing standard and lightweight balls. *Journal of Physical Therapy Science*, 32:816–822, 12 2020.

[81] Y. S. Yoon and J. M. Mansour. The passive elastic moment at the hip. *Journal of Biomechanics*, 15(12):905–910, 1982.

[82] Y. Yuan, S.-E. Wei, T. Simon, K. Kitani, and J. Saragih. Simpoe: Simulated character control for 3d human pose estimation. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2021.

[83] S. Zhang, B. L. Bhatnagar, Y. Xu, A. Winkler, P. Kadlecek, S. Tang, and F. Bogo. Rohm: Robust human motion reconstruction via diffusion. In *CVPR*, 2024.

[84] W. Zhang, M. Zhu, and K. G. Derpanis. From actemes to action: A strongly-supervised representation for detailed action understanding. In *2013 IEEE International Conference on Computer Vision*, pages 2248–2255, 2013.

[85] Y. Zhang, J. O. Kephart, Z. Cui, and Q. Ji. Physpt: Physics-aware pretrained transformer for estimating human dynamics from monocular videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2305–2317, June 2024.

[86] C. Zheng, S. Zhu, M. Mendieta, T. Yang, C. Chen, and Z. Ding. 3D Human Pose Estimation with Spatial and Temporal Transformers. In *Proceedings of the IEEE International Conference on Computer Vision*, 2021.

[87] X. Zhou, Q. Huang, X. Sun, X. Xue, and Y. Wei. Towards 3D Human Pose Estimation in the Wild: A Weakly-Supervised Approach. In *Proceedings of the IEEE International Conference on Computer Vision*, volume 2017-October, 2017.

[88] Y. Zhou, C. Barnes, J. Lu, J. Yang, and H. Li. On the continuity of rotation representations in neural networks. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2019.

[89] W. Zhu, X. Ma, Z. Liu, L. Liu, W. Wu, and Y. Wang. Learning human motion representations: A unified perspective. In *Proceedings of the International Conference on Computer Vision*, 2023.