Muscles in Time: Learning to Understand Human Motion by Simulating Muscle Activations

David Schneider[*,†] **Simon Reiß**[*] **Marco Kugler**[*] **Alexander Jaus**[*] **Kunyu Peng**[*]

Susanne Sutschet[*] **M. Saquib Sarfraz**[‡] **Sven Matthiesen**[*] **Rainer Stiefelhagen**[*]

Abstract

Exploring the intricate dynamics between muscular and skeletal structures is pivotal for understanding human motion. This domain presents substantial challenges, primarily attributed to the intensive resources required for acquiring ground truth muscle activation data, resulting in a scarcity of datasets. In this work, we address this issue by establishing *Muscles in Time (MinT)*, a large-scale synthetic muscle activation dataset. For the creation of *MinT*, we enriched existing motion capture datasets by incorporating muscle activation simulations derived from biomechanical human body models using the OpenSim platform, a common approach in biomechanics and human motion research. Starting from simple pose sequences, our pipeline enables us to extract detailed information about the timing of muscle activations within the human musculoskeletal system. *Muscles in Time* contains over nine hours of simulation data covering 227 subjects and 402 simulated muscle strands. We demonstrate the utility of this dataset by presenting results on neural network-based muscle activation estimation from human pose sequences with two different sequence-to-sequence architectures.

1 Introduction

Like prisoners in Plato's cave, neural networks for human motion understanding often rely on indirect representations rather than direct, biologically grounded data. In Plato's allegory, prisoners in a cave see only shadows cast on the wall, not the true objects. Similarly, neural networks trained on accessible data, such as RGB and depth-based video recordings or motion capture, only perceive surface-level appearance of motion in contrast to the inner mechanics of the human body.

This reliance on external visual observations provides an incomplete understanding of the true complexities of human motion. Just as the prisoners lack a direct view of the objects casting the shadows, current models lack exposure to the internal workings of the human body, such as the muscle activations driving motion. This gap limits their ability to develop an in-depth understanding of physical exertion, motion difficulty, and mass impact on the body.

Our community has progressed from capturing human motion with camera sensors and predicting activities to pose-based recognition systems that account for the body and its motion over time. These advances, while significant, still overlook the interplay of muscle activations, which are the root of pose sequences and patterns.

Collecting electromyographic (EMG) data or more commonly used surface electromyographic (sEMG) data, as a measure of muscle activation, presents challenges. It is resource intensive, requiring specialized equipment, controlled environments, and is an invasive procedure. Existing

[*]Karlsruhe Institute of Technology
[†]Corresponding author: david.schneider@kit.edu
[‡]Mercedes-Benz Tech Innovation

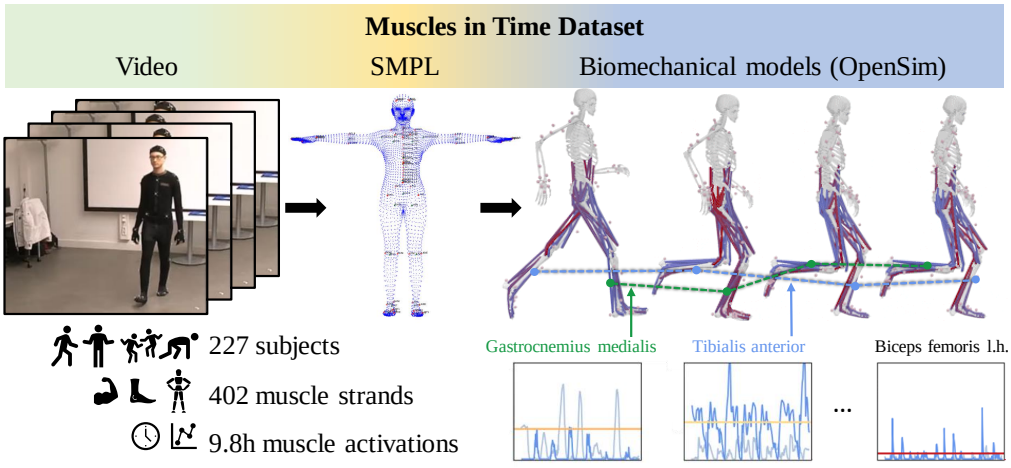

Figure 1: Simulation pipeline of the Muscles in Time dataset. The SMPL representation is extracted from videos, then, the SMPL represented motions are mapped to bio-mechanically validated human body models to simulate fine-grained muscle activation, connecting computer vision with biomechanical research. Bottom right: two activation sequences for exemplary muscles. Images from [40, 13]

EMG and sEMG datasets are small, limited in scope, and not representative of the variety of human motions. These limitations hinder the development of neural networks that can generalize across different types of motion and subjects.

While acknowledging the contributions of EMG and sEMG datasets, we identify an opportunity to supplement this domain with a synthetic dataset that overcomes some limitations of real-world data collection. The strength of our dataset lies in its scale and detail of muscle activation data, a feat not achievable through conventional methods alone.

Every dataset, simulated or real, has domain-specific fidelity and relevance. Real-world recordings offer authenticity that underpins our understanding of human biomechanics with nuances, such as EMG measurements being subject-specific and varying over the course of one day. Simulated datasets, like ours, offer a complementary perspective by providing comprehensive data for the understanding of muscle activation patterns through a scalable data acquisition pipeline.

In this work, we present a comprehensive large-scale dataset incorporating muscle activation information. We enrich existing motion capture datasets with muscle activation simulations from biomechanical models of the human body. Our pipeline uses simple pose and shape sequences with estimated weight and mass of the human body to simulate muscle activations for individual movements. Using this, we generate the muscles' activation that fit the provided human motions. Figure 1 provides an overview of our pipeline.

We showcase the utility of muscle activations as an additional data type for human motion understanding and gather insights by visualizing the intricate details of our data. Our dataset, the first of its magnitude and detail, describes muscle activation across a wide array of movements. By enhancing the current set of tools available to researchers, we expand the potential for scientific investigation and innovation in the study of human motion.

2 Related Work

Human Motion Analysis Datasets EMG based Muscle activation analysis is a well-established field in biomechanical research. Still, publicly available databases including experimentally measured muscle excitation using sEMG are often small in size or cover a small range of muscles or motion variations [19, 25, 71, 22, 44, 39, 54, 35, 28, 43, 58]. The dataset proposed by Zhang *et al*. [71] contains 5 persons and leveraged 8 EMG sensors. The KIMHu dataset [25], for example, includes sEMG data of four upper limb muscles measured during different arm exercises performed by 20 subjects. The MIA Dataset [9] includes sEMG signals for eight muscles in total (upper and lower limb) across 10 subjects who performed 15 different exercises, e.g., running, jumping jacks, squats,

Table 1: Comparison between recent muscle activation datasets and *Muscles in Time*.

	Year	Subjects	Act. Vals.	Duration	Actions	GRF	RGB	Depth	Activation	Skeleton	Description
Camargo *et al.* [4]	2021	22	11	10 min	4	×	×	×	✓	×	×
Feldotto *et al.* [19]	2022	5	7	10 min	4	×	×	×	✓	×	×
KIMHu [25]	2023	20	4	10 h	3	✓	✓	✓	✓	✓	×
MuscleMap [48]	2023	N/A	20[a]	∼25 h[b]	135	✓	✓	✓	×	✓	×
MiA [9]	2023	10	8	12.5 h	15	✓	✓	×	×	✓	×
MinT (ours)	2024	227	402[c,d]	10 h	187	✓[d]	✓[d,e]	✓[d]	✓[d]	✓	✓

[a] Clip-wise binary labels. [b] Coarse estimation based on 15,004 clips of 3-9s. [c] Muscle strands, some muscles represented by multiple strands. [d] Simulated data. [e] From [57]

and elbow punches. MuscleMap is a video-based muscle activation estimation dataset, which assigns binary muscle activation labels to action categories, involving 20 muscle groups and 135 actions [48]. OpenSim is an open-source software platform for musculoskeletal modeling, simulation, and analysis. It is used in various research areas such as biomechanics research, orthopedics and rehabilitation science, and medical device design [14, 59]. The state-of-the-art process in OpenSim for simulating muscle activations of a certain task requires subject-specific motion and force data. In most cases, those are measured in experimental studies which are time-consuming and extensive.

Skeleton-based Vision Models Skeleton-based action recognition [20, 1] is pivotal in decoding human actions from video footage, providing a streamlined and insightful depiction of human poses and movements that remains invariant to changes in appearance, illumination, and backdrop. This approach enhances the identification of dynamic skeletal characteristics essential for precise action recognition, finding utility across surveillance, human-computer interaction, and medical fields. The goal of skeleton-based action recognition is to classify actions based on skeletal geometry information [29, 37, 42, 17, 49, 66, 47, 64, 6]. Predominantly, the techniques employed are based on graph convolutional neural networks (GCN)[31, 68, 61, 8, 69, 7], with newer methods adopting transformer architectures [62, 51, 34, 73, 15, 65]. Chen *et al.* [7] proposed channel-wise topology refinement graph convolution for skeleton-based action recognition. Yan *et al.* [67] proposed skeleton masked auto encoder to achieve skeleton sequence pretraining which delivers promising benefits for the skeleton based action recognition. Apart from the GCN and transformer based models, PoseC3D is proposed by Duan *et al.* [17] to use 3D convolutional neural networks on the heat map figures painted by the skeleton joints.

Sequence-to-sequence Models Sequence to sequence models [45, 30, 10, 72, 36, 60, 21] are a class of deep neural network architectures designed to transform sequences from one domain into sequences in another domain, typically used in applications such as machine translation, speech recognition, and text summarization. These models generally consist of an encoder that processes the input sequence and a decoder that generates the output sequence, facilitating the learning of complex sequence mappings through recurrent neural networks (RNNs) [38, 46, 50, 27] or transformer-based architectures [16, 26]. Chan *et al*. [5] proposed Imputer method by using imputation and dynamic programming to achieve sequence modelling. Colombo *et al*. [12] used guiding attention for sequence-to-sequence modelling for dialogue activities prediction. Rae *et al*. [53] proposed compressed transformer architecture for long-range sequence modelling. Foo *et al*. [21] proposed a unified pose sequence modelling method for human behavior understanding.

3 The Muscles In Time dataset

To develop the *Muscles in Time (MinT)* framework, we harnessed the comprehensive AMASS dataset, which consolidates various marker-based motion capture (mocap) sequences into a uniform representation using the MoSh++ method, resulting in Skinned Multi-Person Linear Model (SMPL) parametric representations for body pose and shape. AMASS amalgamates mocap data from multiple sources, including the KIT Whole-Body Human Motion Database [41], BMLrub, and BMLmovi [23],

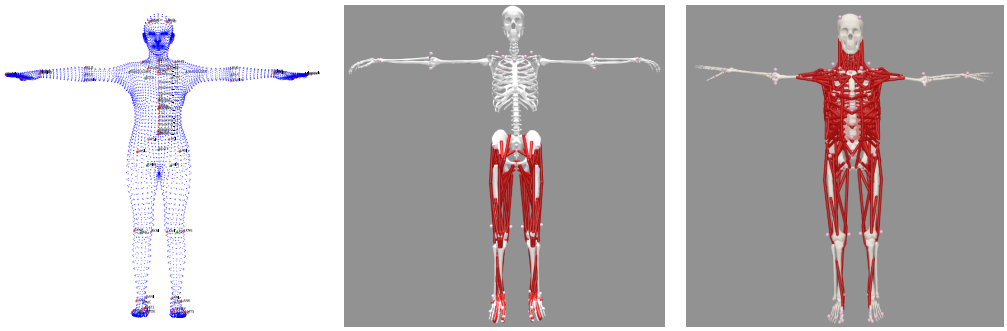

Figure 2: The AMASS body model with specific indices mapped onto the OpenSim lower body model by Lai *et al*. [33] (middle) and model of the thoracolumbar region by Bruno *et al*. [3] (right). Best viewed by zooming in.

encompassing over 11,000 motion captures from more than 300 subjects. This extensive collection enables the analysis of a broad array of human movements, providing a rich basis for studying diverse motion patterns.

The SMPL model serves as a pivotal link, translating mocap data from AMASS into mesh representations which we use to transfer the data into a format compatible with the OpenSim [13] platform. OpenSim is instrumental in constructing intricate biomechanical models that simulate the musculoskeletal system's physical and mechanical properties, allowing for an in-depth analysis of human motion. These models are intricate, requiring precise definitions of joints, masses, inertia, and muscle parameters, such as maximum isometric force, which act as the force-generating actuators.

In this work, we abstain from developing new biomechanical models due to the complexity and expertise required. Instead, we utilize established, pre-validated models, specifically the lower body model by Lai *et al*. [33] and the thoracolumbar region body model by Bruno *et al*. [3], see Figure 2. These models simulate muscle activations for an extensive network of individual muscle strands across various muscle groups, providing a comprehensive simulation of human musculature. A detailed list of these muscle groups and their function in the human body is provided in the Appendix.

Tailoring body model parameters to an individual's anatomical properties results in similar difficulties as with the creation of new body models, therefore parameters are commonly used as specified in the validated original models [33, 3], in the OpenSim community. We follow this approach, providing simulation results for standard models rather than subject specific human bodies.

To integrate human motion data from AMASS with OpenSim, we map virtual mocap markers to the SMPL-H body mesh's surface vertices, following the method proposed by Bittner *et al*. [2]. This results in a selection of 67 strategically placed vertices that represent marker positions on the body mesh, visualized on the left of Figure 2. We deliberately exclude soft tissue dynamics from the SMPL-H mesh generation to maintain consistent marker positions during motion.

Despite OpenSim's automatic scaling capabilities, manual adjustments of marker positions are sometimes necessary to reconcile differences between simulated and real-world data. These adjustments are made on a subject-specific basis, rendering our pipeline semi-automatic. The manually adjusted marker positions are documented and shared to ensure the reproducibility of our simulations.

AMASS lacks data on external ground reaction forces or contact forces, which are crucial for realistic motion simulation. To address this, we integrate the OpenSimAD [18] implementation used in the OpenCap [63] project, which calculates ground reaction forces based on kinematic data and the musculoskeletal model. We employ a tailored parameter setup to optimize the trajectory problem, balancing computational load and accuracy.

Kinematic data is analyzed using OpenSim's *Inverse Kinematics* method. Muscle activations for the lower body are derived from a trajectory optimization problem descibed in [63]. The estimated ground reaction forces from this problem serve as inputs for the *Static Optimization* method, which calculates muscle activations for the thoracolumbar region.

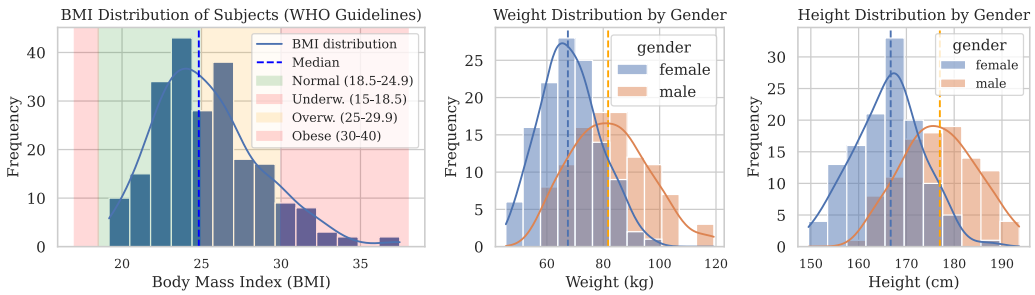

Figure 3: Approximated weight and height distribution of the analysed subjects in the MinT dataset.

Due to the computational demands of the trajectory optimization problem, we process the data in segments, ensuring manageable computation times without compromising the continuity of the motion capture sequences. We implement overlapping buffers to mitigate inaccuracies during segment processing, discarding data that fails to meet our stringent error tolerance criteria to maintain a high standard of data quality. Further details on implementation and design decisions of our simulation process are presented in the Appendix.

The Muscles in Time (MinT) dataset represents a significant contribution to the field of biomechanical and computer vision simulation. By integrating and refining existing methodologies, we present a robust pipeline that facilitates the accurate simulation of human muscles in motion by combining established biomechanical models with high quality mocap data. To ensure reproducibility we will release all relevant data and details of our simulation process to the scientific community.

3.1 Dataset Composition

Due to missing information of external forces based on object interactions, inaccurate motion capture recordings or non-converging simulations, the *MinT* dataset covers a subset of its originating datasets in AMASS and does not follow their respective dataset statistics.

Anthropometrics While the motion capture recordings in AMASS provide gender labels, information about subjects height and weight is approximated from the SMPL body model. Body weight is calculated by volume resulting from average shape parameters, which follows the approach of Bittner *et al*. [2]. The weight is relevant for the calculation of ground reaction forces and the distribution of weight in the model, affecting the muscle activation in different parts of the body.

The Figure 3 showcases the distribution of weight, indicating significant diversity. Underweight subjects are slightly underrepresented in the dataset, subjects on the obese range are well represented.

Composition of Subdatasets Within AMASS, *MinT* is limited to the subdatasets EyesJapan, BMLrub, KIT, BMLmovi, and TotalCapture. Figure 9 in the appendix shows the ratio of the originating subdatasets in our final simulation results as well as the average sequence length within these subdatasets. The short sequences in BMLmovi typically depict single activities, while the longer ones for example in JapanEyes capture a more diverse range of motions within a single sequence. Since we compute activation information for shorter segments and rejoin them afterwards, longer sequences are more prone to gaps in the analysis due to individually failing segment computations.

Motion Diversity Figure 4 displays the frequencies of grouped activities on a logarithmic scale. The action labels are based on the BABEL dataset, a large annotation dataset which is coupled with AMASS. Most interesting are the dynamic actions as well as movement types, since expected muscle activations for simple dynamic actions are well documented and we present a short qualitative analysis based on this in Section 3.2.

3.2 Data Analysis, Validation, and Visualization

In Figure 5 (left) we explore the interrelation between different activities by investigating our simulated muscle activation time-series. To this end, we extract features from the temporal muscle activation sequences using tsfresh [11], a commonly used framework in time series analysis that

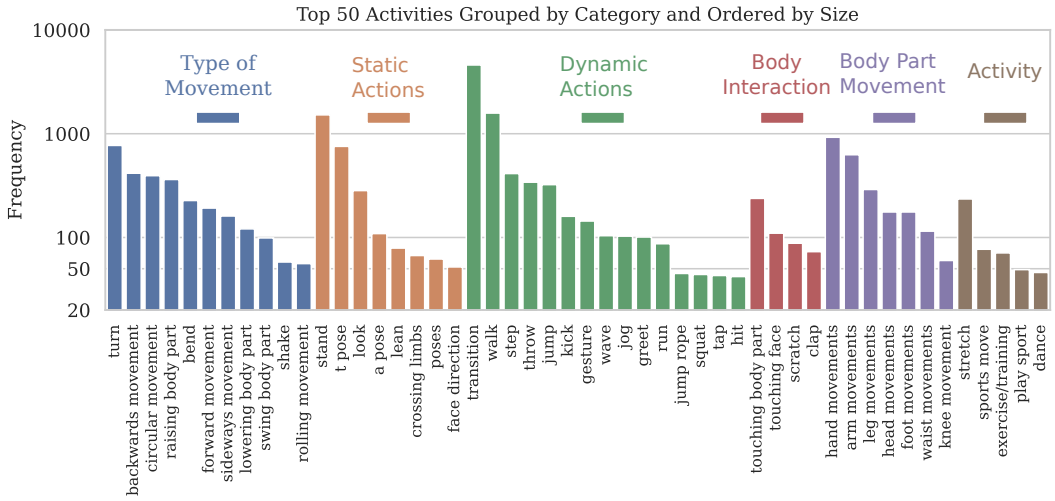

Figure 4: Prevalence of different motions in the MinT dataset.

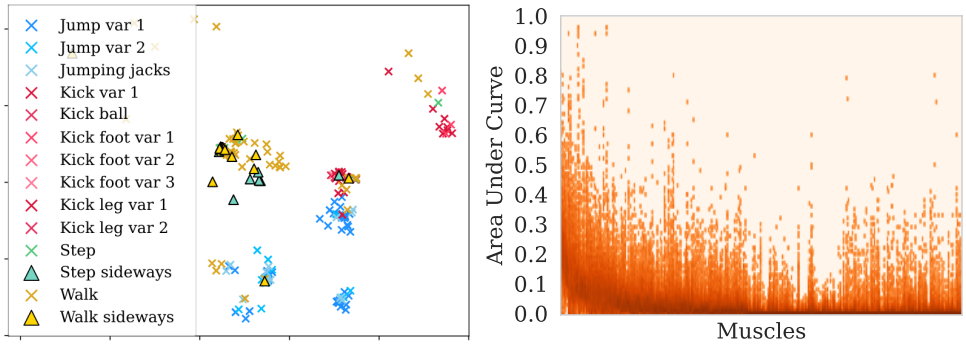

Figure 5: **Left:** Clustering of multiple activities within the BMLmovi dataset by muscle activation features. **Right:** Column wise color coded histograms of areas under muscle activation curves for 402 muscles strains, sorted by histogram medians. Log-normalized color map, best displayed in color.

extracts a feature vector based on time series characteristics such as mean, skewness, standard deviation *etc*. We choose distinct and descriptive groups of activities from the BMLmovi subset such as jumping, kicking, stepping and walking, the resulting features were normalized and clustered using FINCH [55] and visualized with h-NNE [56]. It can be observed, that activities do not only cluster together based on variations within the same category (e.g., different types of jumps, including jumping jacks), but also align closely across different categories, when they share similar motion patterns (e.g. sideways movements). This underlines the descriptive information contained in our simulated muscle activation sequences for characterizing activities.

4 Motion to Muscle Activation Estimation Benchmark

While OpenSim provides a means for simulating muscle activations, it is both highly compute intensive as well as sensitive regarding hyper parameters as described in Section 6. These properties limit it to be used by experts in an offline manner and prevent usage in everyday applications. In this section we explore the usage of MINT as a training dataset for the estimation of muscle activation based on pose motion. Such networks provide muscle activation estimation in an instant and can easily be deployed for various downstream tasks.

Given pose motion sequences we use the preprocessing step defined by [24] which adjusts skeletal structure to a uniform format and normalizes positions and enriches the resulting data points with

additional features. This procedure maps each input to a 263-dim descriptor resulting in samples of the form $x = [x_1, ..., x_T]$, $x_t \in \mathbb{R}^{263}$. For training our models we segment the resulting data into clips of 1.4 second sampled at 20 frames per second, resulting in $T = 28$ input frames. Given a network $f_\Theta : \mathbb{R}^{T \times d} \mapsto \mathbb{R}^{T \times m}$ we predict $f(x) = y$ with $y = [y_1, ..., y_T]$, $m = 402$ being the number of individual muscle strain activations simulated in our dataset, consisting of 80 lower body muscle strains from [18] and 322 muscle strains for the upper thoracolumbar region body model [3]. Evaluation is performed by calculating Root Mean Squared Error (RMSE), Pearson Correlation Coefficient (PCC), and Symmetric Mean Absolute Percentage Error (SMAPE). RMSE is commonly used but highly susceptible to data scaling, resulting in significantly lower error values for downscaled data. In practice, EMG signals vary strongly between subjects, scaling of signals is therefore a common preprocessing step. PCC is a good indicator for muscle activation series similarity, since it is scale and offset invariant. SMAPE allows for considering fixed offsets as error while being less sensitive to scaling in comparison to RMSE. PCC and SMAPE are calculated for each muscle strain individually and averaged. For our benchmark we use the train, val and test splits defined by the BABEL dataset [52]. Evaluation results are reported separately for muscles of the upper and lower body model.

5 Experiments

We evaluate two different architectures on Muscles In Time. Since we make use of human motion as input for our predictor, we adapted a common architecture for motion-to-motion prediction from [70] to the task of motion-to-muscle activation prediction by simply exchanging its prediction head. We further evaluate a simple transformer architecture with 4, 8 or 16 transformer layers, results for the lower and upper body model are listed in Table 2. All models are trained from scratch for 300k iterations with a batch size of 256 unless noted otherwise. More details on the model implementations can be found in the supplementary.

Table 2: Human motion-to-muscle activation prediction results for the lower- and upper body model.

Motion	Transformer 4 Layer			Transformer 8 Layer			Transformer 16 Layer			VQ-VAE [70]		
	RMSE↓	PCC↑	SMAPE↓	RMSE↓	PCC↑	SMAPE↓	RMSE↓	PCC↑	SMAPE↓	RMSE↓	PCC↑	SMAPE↓
Lower body model												
overall	0.049	0.52	52.8	**0.048**	0.53	49.1	**0.048**	**0.54**	**45.1**	0.058	0.40	59.7
jump	0.051	0.69	57.2	**0.049**	**0.71**	54.2	0.051	**0.71**	**52.3**	0.062	0.52	66.7
kick	0.055	0.59	62.0	0.054	0.60	59.2	**0.053**	**0.62**	**54.8**	0.069	0.38	69.1
stand	0.047	0.55	52.7	**0.046**	0.56	48.8	**0.046**	**0.58**	**45.0**	0.056	0.42	60.0
walk	0.046	0.74	50.5	**0.044**	0.76	46.2	**0.044**	**0.77**	**42.4**	0.053	0.66	57.7
jog	0.048	0.69	57.5	**0.046**	0.70	55.2	**0.046**	**0.71**	**51.1**	0.059	0.58	64.8
dance	0.059	0.64	64.2	**0.057**	0.64	61.3	**0.057**	**0.65**	**58.5**	0.070	0.40	71.4
Upper body model												
overall	0.033	0.53	109.3	**0.032**	0.54	108.4	0.033	**0.55**	**107.7**	0.041	0.32	115.3
jump	0.052	0.61	114.0	**0.051**	0.62	112.8	0.052	**0.63**	**112.7**	0.064	0.38	118.1
kick	0.046	0.62	116.4	0.046	0.62	115.7	**0.044**	**0.65**	**114.8**	0.058	0.35	122.2
stand	0.030	0.53	112.2	0.029	0.54	111.4	**0.028**	**0.55**	**110.5**	0.039	0.34	117.6
walk	**0.019**	0.61	103.9	**0.019**	0.62	103.3	**0.019**	**0.63**	**102.6**	0.028	0.43	110.2
jog	0.029	0.69	111.8	**0.028**	0.70	111.2	0.029	**0.71**	**110.8**	0.040	0.52	117.1
dance	0.040	0.57	123.4	**0.034**	0.58	**121.7**	0.036	**0.59**	121.8	0.046	0.29	127.3

The evaluated transformer architecture improved over the adapted VQ-VAE model in all metrics on all evaluated motion types, with the 16-layer transformer generally showing best results, indicating that even larger models can be trained with improvement on our dataset. The results of the experiment also show the importance of reporting PCC and SMAPE, since the differences on RMSE are marginal while PCC shows significant improvements as does SMAPE on larger models. We suspect this to be the case, since many muscles in the human body are mostly relatively inactive unless required for specific motions. For a simple analysis of this effect, we calculated the integral for each individual ground truth muscle activation sequence in all our validation set chunks and created 402 color coded

histograms which are sorted by median and vertically displayed side by side on the right hand side of Figure 5 (one column in the image is a single muscle activation integral area frequency histogram). A wide range of muscles are rarely activated, resulting in the majority of activation sequences displaying integral areas significantly below 0.1 or 0.05. This property is challenging for RMSE and SMAPE, average RMSE reports a small error, since most activations are close to zero and SMAPE reports a high percentage error, since a deviation from a close to zero value is more likely to result in a high percentage deviation. For similar reasons, the upper body model displays lower RMSE and higher SMAPE, the upper body model contains a larger number of small and rarely activated muscles in contrast to the lower body model.

To provide a more detailed analysis we list the results on the collection of all available muscle strains in the main paper, but list further evaluations on carefully chosen subsets of major motion inducing body muscles in the appendix. We recommend future users of our dataset to consider actively evaluating on either the full range of provided muscle activations or chosing one of these muscle strand subsets depending on their specific application. Please also see the appendix for additional experiments as well as a comparison to the work of [9].

5.1 Qualitative Results

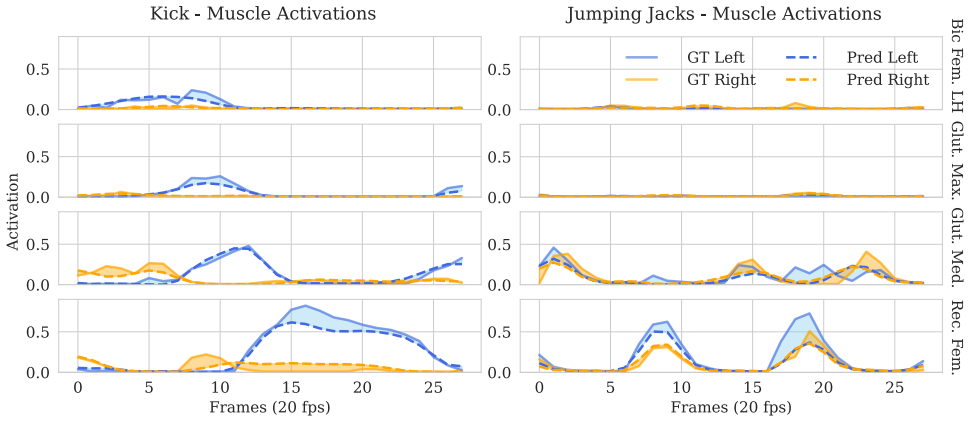

Figure 6: Example lower body muscle activations (split in left and right muscle strands) for the actions *kick* and *jumping jacks*. It is clearly visible that the kick is performed with the left leg. During *jumping jacks*, *glueteus medius* and *rectus femoris* are activated alternatingly for both legs.

In Figure 6 we list two examples from our dataset, one displaying the action *kick*, the other displaying the action *jumping jacks*, predictions are calculated with the 8-layer transformer architecture. The figure displays four key muscles essential for lower body locomotion; *biceps femoris long head* (knee flexion and hip extension), *gluteus maximus* (hip extension and external rotation), *gluteus medius* (abduction and medial rotation of the hip), and *rectus femoris* (hip flexion and knee extension), each for the left and right body half. The kick is clearly executed with the left leg with *rectus femoris* providing the force for the swing in the second half of the motion and the other muscles of the left leg preparing it in the first half. During *jumping jacks*, *gluteus medius* and *rectus femoris* are activated alternatingly for both legs. Predicted muscle activations closely follow the ground truth from our dataset, with slight underestimation at the activation peaks. Similar estimation quality can be observed across the test set and we refer the reader to the appendix where we provide a larger number of randomly selected results for qualitative analysis.

6 Discussion

We believe, that enhancing models through detailed muscle activation data aligned with human motion is a worthwhile direction to explore in the future, which is now made possible by the presented *MinT* dataset. The dataset offers a large amount of intricately simulated data, based on real human motions, and utilizing bio-mechanically validated musculoskeletal models. By showing that neural models can learn to connect motion input to muscle activation sequences, we broaden the pathway towards models which understand the nuanced interplay between motion and muscles.

Societal Impact While the dataset has a good balance in terms of gender distribution, ethnicity is not distributed equally, and some body-weight types are less represented, impacting the dataset diversity.

Limitations *MinT* is a simulation dataset and while we carefully designed our pipeline and analyze the resulting data, a synthetic-to-real domain gap is unavoidable. In the construction of the dataset, some design choices had to increase simulation robustnees, including safeguards for the vertebral joints against aberrant movements by constraining the range of motion to their natural degrees of freedom. Further, the MinT dataset is limited to motions incorporating feet-ground contact, motions involving the contact of other body parts or other objects than the feet and the ground were excluded. A detailed discussion of these exclusions and the design choices made can be found in the appendix.

7 Conclusion

The quest to analyze human motion necessitates a critical component that has been notably absent: a comprehensive biomechanical dataset. Our contribution, the Muscles in Time (MinT) dataset, addresses this gap by providing an unprecedented collection of synthetic muscle activation data. This dataset encompasses 402 distinct simulated muscle strains, all derived from authentic human movements, thus offering a vital resource for human motion research. Our methodology entails a scalable pipeline that utilizes cutting-edge musculoskeletal models to derive muscle activations from recorded human motion sequences. The culmination of this process is the MinT dataset, which also contains 9.8 hours of time series data representing muscle activations. We demonstrate that neural networks can effectively utilize this muscle activation data to discern patterns linking motion to muscle activation. This represents a significant stride towards a deeper comprehension of human motion from a biomechanical standpoint. The MinT dataset enables the research community in exploration of human motion and muscular dynamics through a data-centric approach. Our work not only enriches the field of biomechanical studies but also paves the way for future advancements in understanding the complex interplay of muscles in human movement.

Acknowledgemets This work has been supported by the Carl Zeiss Foundation through the JuBot project as well as by funding from the pilot program Core-Informatics of the Helmholtz Association (HGF). The authors acknowledge support by the state of Baden-Württemberg through bwHPC. Experiments were performed on the HoreKa supercomputer funded by the Ministry of Science, Research and the Arts Baden-Württemberg and by the Federal Ministry of Education and Research.

References

[1] Ahmad, T., Jin, L., Zhang, X., Lai, S., Tang, G., Lin, L.: Graph convolutional neural network for human action recognition: A comprehensive survey. TAI (2021)

[2] Bittner, M., Yang, W.T., Zhang, X., Seth, A., van Gemert, J., van der Helm, F.C.T.: Towards single camera human 3d-kinematics **23**(1), 341. https://doi.org/10.3390/s23010341, `https://www.mdpi.com/1424-8220/23/1/341`

[3] Bruno, A.G., Bouxsein, M.L., Anderson, D.E.: Development and validation of a musculoskeletal model of the fully articulated thoracolumbar spine and rib cage **137**(8), 081003. https://doi.org/10.1115/1.4030408

[4] Camargo, J., Ramanathan, A., Flanagan, W., Young, A.: A comprehensive, open-source dataset of lower limb biomechanics in multiple conditions of stairs, ramps, and level-ground ambulation and transitions. Journal of Biomechanics **119**, 110320 (2021)

[5] Chan, W., Saharia, C., Hinton, G.E., Norouzi, M., Jaitly, N.: Imputer: Sequence modelling via imputation and dynamic programming. In: ICML (2020)

[6] Chen, Y., Peng, K., Roitberg, A., Schneider, D., Zhang, J., Zheng, J., Liu, R., Chen, Y., Yang, K., Stiefelhagen, R.: Unveiling the hidden realm: Self-supervised skeleton-based action recognition in occluded environments. arXiv preprint arXiv:2309.12029 (2023)

[7] Chen, Y., Zhang, Z., Yuan, C., Li, B., Deng, Y., Hu, W.: Channel-wise topology refinement graph convolution for skeleton-based action recognition. In: ICCV (2021)

[8] Cheng, K., Zhang, Y., Cao, C., Shi, L., Cheng, J., Lu, H.: Decoupling gcn with dropgraph module for skeleton-based action recognition. In: ECCV (2020)

[9] Chiquier, M., Vondrick, C.: Muscles in action. In: CVPR (2023)

[10] Chiu, C.C., Sainath, T.N., Wu, Y., Prabhavalkar, R., Nguyen, P., Chen, Z., Kannan, A., Weiss, R.J., Rao, K., Gonina, E., et al.: State-of-the-art speech recognition with sequence-to-sequence models. In: ICASSP (2018)

[11] Christ, M., Braun, N., Neuffer, J., Kempa-Liehr, A.W.: Time series feature extraction on basis of scalable hypothesis tests (tsfresh–a python package). Neurocomputing **307**, 72–77 (2018)

[12] Colombo, P., Chapuis, E., Manica, M., Vignon, E., Varni, G., Clavel, C.: Guiding attention in sequence-to-sequence models for dialogue act prediction. In: AAAI (2020)

[13] Delp, S.L., Anderson, F.C., Arnold, A.S., Loan, P., Habib, A., John, C.T., Guendelman, E., Thelen, D.G.: OpenSim: Open-source software to create and analyze dynamic simulations of movement **54**(11), 1940–1950. https://doi.org/10.1109/TBME.2007.901024, `https://ieeexplore.ieee.org/abstract/document/4352056`

[14] Delp, S.L., Anderson, F.C., Arnold, A.S., Loan, P., Habib, A., John, C.T., Guendelman, E., Thelen, D.G.: Opensim: open-source software to create and analyze dynamic simulations of movement. IEEE transactions on bio-medical engineering **54**(11), 1940–1950 (2007). https://doi.org/10.1109/TBME.2007.901024

[15] Ding, K., Liang, A.J., Perozzi, B., Chen, T., Wang, R., Hong, L., Chi, E.H., Liu, H., Cheng, D.Z.: HyperFormer: Learning expressive sparse feature representations via hypergraph transformer. In: SIGIR (2023)

[16] Dong, L., Xu, S., Xu, B.: Speech-transformer: a no-recurrence sequence-to-sequence model for speech recognition. In: ICASSP (2018)

[17] Duan, H., Zhao, Y., Chen, K., Lin, D., Dai, B.: Revisiting skeleton-based action recognition. In: CVPR (2022)

[18] Falisse, A., Serrancoli, G., Dembia, C.L., Gillis, J., Groote, F.D.: Algorithmic differentiation improves the computational efficiency of OpenSim-based trajectory optimization of human movement **14**(10), e0217730. https://doi.org/10.1371/journal.pone.0217730, `https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0217730`

[19] Feldotto, B., Soare, C., Knoll, A., Sriya, P., Astill, S., de Kamps, M., Chakrabarty, S.: Evaluating muscle synergies with emg data and physics simulation in the neurorobotics platform. Frontiers in Neurorobotics (2022)

[20] Feng, L., Zhao, Y., Zhao, W., Tang, J.: A comparative review of graph convolutional networks for human skeleton-based action recognition. Artificial Intelligence Review (2022)

[21] Foo, L.G., Li, T., Rahmani, H., Ke, Q., Liu, J.: Unified pose sequence modeling. In: CVPR (2023)

[22] Furmanek, M.P., Mangalam, M., Yarossi, M., Lockwood, K., Tunik, E.: A kinematic and emg dataset of online adjustment of reach-to-grasp movements to visual perturbations. Scientific data **9**(1), 23 (2022)

[23] Ghorbani, S., Mahdaviani, K., Thaler, A., Kording, K., Cook, D., Blohm, G., Troje, N.: Movi: A large multipurpose motion and video dataset. arxiv 2020. arXiv preprint arXiv:2003.01888 (2020)

[24] Guo, C., Zou, S., Zuo, X., Wang, S., Ji, W., Li, X., Cheng, L.: Generating diverse and natural 3d human motions from text. In: CVPR (2022)

[25] Hernández, Ó.G., Lopez-Castellanos, J.M., Jara, C.A., Garcia, G.J., Ubeda, A., Morell-Gimenez, V., Gomez-Donoso, F.: A kinematic, imaging and electromyography dataset for human muscular manipulability index prediction. Scientific Data (2023)

[26] Hrinchuk, O., Popova, M., Ginsburg, B.: Correction of automatic speech recognition with transformer sequence-to-sequence model. In: ICASSP (2020)

[27] Jaitly, N., Le, Q.V., Vinyals, O., Sutskever, I., Sussillo, D., Bengio, S.: An online sequence-to-sequence model using partial conditioning. NeurIPS (2016)

[28] Jarque-Bou, N.J., Vergara, M., Sancho-Bru, J.L., Gracia-Ibáñez, V., Roda-Sales, A.: A calibrated database of kinematics and emg of the forearm and hand during activities of daily living. Scientific data **6**(1), 270 (2019)

[29] Ke, Q., Bennamoun, M., An, S., Sohel, F., Boussaid, F.: A new representation of skeleton sequences for 3d action recognition. In: ICCV (2017)

[30] Keneshloo, Y., Shi, T., Ramakrishnan, N., Reddy, C.K.: Deep reinforcement learning for sequence-to-sequence models. TNNLS (2019)

[31] Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907 (2016)

[32] Kocabas, M., Athanasiou, N., Black, M.J.: Vibe: Video inference for human body pose and shape estimation. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2020)

[33] Lai, A.K., Arnold, A.S., Wakeling, J.M.: Why are antagonist muscles co-activated in my simulation? A musculoskeletal model for analysing human locomotor tasks. Annals of Biomedical Engineering **45**, 2762–2774 (2017)

[34] Lee, J., Lee, M., Lee, D., Lee, S.: Hierarchically decomposed graph convolutional networks for skeleton-based action recognition. arXiv preprint arXiv:2208.10741 (2022)

[35] Lencioni, T., Carpinella, I., Rabuffetti, M., Marzegan, A., Ferrarin, M.: Human kinematic, kinetic and emg data during different walking and stair ascending and descending tasks. Scientific data **6**(1), 309 (2019)

[36] Li, C., Zhang, Z., Lee, W.S., Lee, G.H.: Convolutional sequence to sequence model for human dynamics. In: CVPR (2018)

[37] Liu, M., Liu, H., Chen, C.: Enhanced skeleton visualization for view invariant human action recognition. PR (2017)

[38] Ma, L., Zhao, Y., Wang, B., Shen, F.: A multi-step sequence-to-sequence model with attention lstm neural networks for industrial soft sensor application. IEEE Sensors Journal (2023)

[39] Malešević, N., Olsson, A., Sager, P., Andersson, E., Cipriani, C., Controzzi, M., Björkman, A., Antfolk, C.: A database of high-density surface electromyogram signals comprising 65 isometric hand gestures. Scientific Data **8**(1), 63 (2021)

[40] Mandery, C., Terlemez, Ö., Do, M., Vahrenkamp, N., Asfour, T.: The kit whole-body human motion database. In: 2015 International Conference on Advanced Robotics (ICAR). pp. 329–336. IEEE (2015)

[41] Mandery, C., Terlemez, O., Do, M., Vahrenkamp, N., Asfour, T.: Unifying representations and large-scale whole-body motion databases for studying human motion. TRO (2016)

[42] Marinov, Z., Schneider, D., Roitberg, A., Stiefelhagen, R.: Multimodal generation of novel action appearances for synthetic-to-real recognition of activities of daily living. In: 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 11320–11327. IEEE (2022)

[43] Matran-Fernandez, A., Rodríguez Martínez, I.J., Poli, R., Cipriani, C., Citi, L.: Seeds, simultaneous recordings of high-density emg and finger joint angles during multiple hand movements. Scientific data **6**(1), 186 (2019)

[44] Moreira, L., Figueiredo, J., Fonseca, P., Vilas-Boas, J.P., Santos, C.P.: Lower limb kinematic, kinetic, and emg data from young healthy humans during walking at controlled speeds. Scientific data **8**(1), 103 (2021)

[45] Neubig, G.: Neural machine translation and sequence-to-sequence models: A tutorial. arXiv preprint arXiv:1703.01619 (2017)

[46] Orvieto, A., Smith, S.L., Gu, A., Fernando, A., Gulcehre, C., Pascanu, R., De, S.: Resurrecting recurrent neural networks for long sequences. arXiv preprint arXiv:2303.06349 (2023)

[47] Peng, K., Roitberg, A., Yang, K., Zhang, J., Stiefelhagen, R.: Delving deep into one-shot skeleton-based action recognition with diverse occlusions. TMM (2023)

[48] Peng, K., Schneider, D., Roitberg, A., Yang, K., Zhang, J., Sarfraz, M.S., Stiefelhagen, R.: Musclemap: Towards video-based activated muscle group estimation. arXiv preprint arXiv:2303.00952 (2023)

[49] Peng, K., Yin, C., Zheng, J., Liu, R., Schneider, D., Zhang, J., Yang, K., Sarfraz, M.S., Stiefelhagen, R., Roitberg, A.: Navigating open set scenarios for skeleton-based action recognition. In: AAAI (2024)

[50] Phan, H., Andreotti, F., Cooray, N., Chén, O.Y., De Vos, M.: Seqsleepnet: end-to-end hierarchical recurrent neural network for sequence-to-sequence automatic sleep staging. TNSRE (2019)

[51] Plizzari, C., Cannici, M., Matteucci, M.: Spatial temporal transformer network for skeleton-based action recognition. In: ICPRW (2021)

[52] Punnakkal, A.R., Chandrasekaran, A., Athanasiou, N., Quiros-Ramirez, A., Black, M.J.: BABEL: Bodies, action and behavior with english labels. In: Proceedings IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR). pp. 722–731 (Jun 2021)

[53] Rae, J.W., Potapenko, A., Jayakumar, S.M., Lillicrap, T.P.: Compressive transformers for long-range sequence modelling. ArXiv **abs/1911.05507** (2019), `https://api.semanticscholar.org/CorpusID:207930593`

[54] Rojas-Martínez, M., Serna, L.Y., Jordanic, M., Marateb, H.R., Merletti, R., Mañanas, M.Á.: High-density surface electromyography signals during isometric contractions of elbow muscles of healthy humans. Scientific data **7**(1), 397 (2020)

[55] Sarfraz, M.S., Sharma, V., Stiefelhagen, R.: Efficient parameter-free clustering using first neighbor relations. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 8934–8943 (2019)

[56] Sarfraz, S., Koulakis, M., Seibold, C., Stiefelhagen, R.: Hierarchical nearest neighbor graph embedding for efficient dimensionality reduction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 336–345 (2022)

[57] Schneider, D., Keller, M., Zhong, Z., Peng, K., Roitberg, A., Beyerer, J., Stiefelhagen, R.: Synthact: Towards generalizable human action recognition based on synthetic data. In: ICRA (2024)

[58] Schreiber, C., Moissenet, F.: A multimodal dataset of human gait at different walking speeds established on injury-free adult participants. Scientific data **6**(1), 111 (2019)

[59] Seth, A., Hicks, J.L., Uchida, T.K., Habib, A., Dembia, C.L., Dunne, J.J., Ong, C.F., DeMers, M.S., Rajagopal, A., Millard, M., Hamner, S.R., Arnold, E.M., Yong, J.R., Lakshmikanth, S.K., Sherman, M.A., Ku, J.P., Delp, S.L.: Opensim: Simulating musculoskeletal dynamics and neuromuscular control to study human and animal movement. PLoS computational biology **14**(7), e1006223 (2018). https://doi.org/10.1371/journal.pcbi.1006223

[60] Shao, L., Gouws, S., Britz, D., Goldie, A., Strope, B., Kurzweil, R.: Generating high-quality and informative conversation responses with sequence-to-sequence models. arXiv preprint arXiv:1701.03185 (2017)

[61] Shi, L., Zhang, Y., Cheng, J., Lu, H.: Two-stream adaptive graph convolutional networks for skeleton-based action recognition. In: CVPR (2019)

[62] Shi, L., Zhang, Y., Cheng, J., Lu, H.: Decoupled spatial-temporal attention network for skeleton-based action-gesture recognition. In: ACCV (2020)

[63] Uhlrich, S.D., Falisse, A., Kidziński, Ł., Muccini, J., Ko, M., Chaudhari, A.S., Hicks, J.L., Delp, S.L.: Opencap: Human movement dynamics from smartphone videos. PLoS computational biology **19**(10), e1011462 (2023)

[64] Wei, Y., Peng, K., Roitberg, A., Zhang, J., Zheng, J., Liu, R., Chen, Y., Yang, K., Stiefelhagen, R.: Elevating skeleton-based action recognition with efficient multi-modality self-supervision. In: ICASSP (2024)

[65] Xin, W., Liu, R., Liu, Y., Chen, Y., Yu, W., Miao, Q.: Transformer for skeleton-based action recognition: A review of recent advances. Neurocomputing (2023)

[66] Xu, Y., Peng, K., Wen, D., Liu, R., Zheng, J., Chen, Y., Zhang, J., Roitberg, A., Yang, K., Stiefelhagen, R.: Skeleton-based human action recognition with noisy labels. arXiv preprint arXiv:2403.09975 (2024)

[67] Yan, H., Liu, Y., Wei, Y., Li, Z., Li, G., Lin, L.: Skeletonmae: graph-based masked autoencoder for skeleton sequence pre-training. In: ICCV (2023)

[68] Yan, S., Xiong, Y., Lin, D.: Spatial temporal graph convolutional networks for skeleton-based action recognition. In: AAAI (2018)

[69] Ye, F., Pu, S., Zhong, Q., Li, C., Xie, D., Tang, H.: Dynamic gcn: Context-enriched topology learning for skeleton-based action recognition. In: MM (2020)

[70] Zhang, J., Zhang, Y., Cun, X., Huang, S., Zhang, Y., Zhao, H., Lu, H., Shen, X.: T2m-gpt: Generating human motion from textual descriptions with discrete representations. In: CVPR (2023)

[71] Zhang, Q.: Experimental data of semg, us imaging, grf, and markers for walking on treadmill across multiple speeds (2021). https://doi.org/10.21227/7beh-f093, `https://dx.doi.org/10.21227/7beh-f093`

[72] Zhong, Z., Schneider, D., Voit, M., Stiefelhagen, R., Beyerer, J.: Anticipative feature fusion transformer for multi-modal action anticipation. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 6068–6077 (2023)

[73] Zhou, Y., Li, C., Cheng, Z.Q., Geng, Y., Xie, X., Keuper, M.: Hypergraph transformer for skeleton-based action recognition. arXiv preprint arXiv:2211.09590 (2022)

[74] Zimmer, P., Appell, H.J.: Funktionelle Anatomie: Grundlagen sportlicher Leistung und Bewegung. Springer Berlin Heidelberg, Berlin, Heidelberg (2021). https://doi.org/10.1007/978-3-662-61482-2

Checklist

The checklist follows the references. Please read the checklist guidelines carefully for information on how to answer these questions. For each question, change the default **[TODO]** to [Yes] , [No] , or [N/A] . You are strongly encouraged to include a **justification to your answer**, either by referencing the appropriate section of your paper or providing a brief inline description. For example:

- Did you include the license to the code and datasets? [Yes] See Appendix.
- Did you include the license to the code and datasets? [No] The code and the data are proprietary.
- Did you include the license to the code and datasets? [N/A]

Please do not modify the questions and only use the provided macros for your answers. Note that the Checklist section does not count towards the page limit. In your paper, please delete this instructions block and only keep the Checklist section heading above along with the questions/answers below.

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes] The dataset is described in Section 3, utility experiments in Section 5.
   (b) Did you describe the limitations of your work? [Yes] See Section 6.
   (c) Did you discuss any potential negative societal impacts of your work? [Yes] See Section 6.
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes] The paper complies with the guidelines.
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [N/A] No theoretical results.
   (b) Did you include complete proofs of all theoretical results? [N/A] No theoretical results.
3. If you ran experiments (e.g. for benchmarks)...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] We provide a link to the code repository in the appendix.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] We define the set of hyperparameters specifying the dataset generation pipeline in Section 3. Hyperparameters chosen to run the time-series prediction experiments are partly defined in Section 5 and partly shown in the appendix. On top, we release the code ensuring reproducibility.
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [No] We experiment with multiple architectures and found that these yielded consistent results. We thus did not execute experiments repeatedly to determine error bars.
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] We dedicate a section in the appendix to discuss the used computational environment and resources.
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [Yes] Yes, the used assets are cited in Section 3
   (b) Did you mention the license of the assets? [Yes] We dedicate a section in the appendix to mention the licenses.
   (c) Did you include any new assets either in the supplemental material or as a URL? [Yes] We provide the proposed *MinT* dataset in this work, available via a URL in the appendix.

(d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [Yes] Generally, we make use of preexisting datasets based on their open license and rely on their discussion of obtaining the consent of the participants which was part of the original publications. On top, we dedicate a section in the appendix to discuss this in more detail.

(e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [Yes] Generally, we make use of preexisting datasets based on their open license and rely on the anonymity or ethics approvals of the original works. On top, we dedicate a section in the appendix to discuss these details.

5. If you used crowdsourcing or conducted research with human subjects...

(a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A] No crowdsourcing or research with human subjects has been done in this work.

(b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A] No crowdsourcing or research with human subjects has been done in this work.

(c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A] No crowdsourcing or research with human subjects has been done in this work.

A Appendix

A.1 Dataset Information

Dataset access and maintenance plan The *MinT* dataset will be provided via the persistent long-term storage service RADAR4KIT (Research Data Repository for KIT), ensuring both uninterrupted and machine readable access. Data published by *RADAR4KIT* is indexed via Metadata following the *Open Archive Initiative* interface which is automatically published to datacite.org and will automatically be referable via a DOI. Data is secured according to *Open Archival Information System* standard ISO 14721:2003 and availability is guaranteed for a minimum of 10 years.

To facilitate the review process and integrate reviewer feedback concerning the data structure (RADAR4KIT data can not be changed easily), we provide an intermediate link for direct download of our data, which will be exchanged with a RADAR4KIT link for the camera ready version.

Currently the dataset can be downloaded under this link (2.2 GB, compressed tar file):
https://s.kit.edu/mint-data

Our code for motion to muscle estimation can be found here:
https://github.com/simplexsigil/motion2muscle.git

License The MinT dataset is build on top of the KIT Whole-Body Human Motion Database, BMLmovi, BMLrub, the EyesJapan dataset and TotalCapture. We make use of AMASS to map from the motions of these original datasets to virtual marker positions in OpenSim.

All of these datasets allow usage of their data for non-commercial scientific research:

- The license of AMASS can be found under https://amass.is.tue.mpg.de/license.html
- The License of BMLmovi and BMLrub can be found under https://www.biomotionlab.ca/movi/
- The KIT Whole-Body Human Motion Database can be used upon citation of the original work as explained here https://download.is.tue.mpg.de/amass/licences/kit.html
- The license for the EyesJapan dataset can be found under http://mocapdata.com/Terms_of_Use.html
- The license for the Total Capture dataset can be found under https://cvssp.org/data/totalcapture/

The Muscles in Time dataset will be published under a CC BY-NC 4.0 license as defined under https://creativecommons.org/licenses/by-nc/4.0/. Researchers making use of this dataset must also agree to the licenses mentioned above which can add additional restrictions depending on the individual sub-dataset.

Our data generation pipeline is licensed under Apache License Version 2.0 as defined under https://apache.org/licenses/LICENSE-2.0.

Code for training our muscle activation estimation networks is licensed under the MIT license as defined under https://opensource.org/license/mit.

Author statement The authors of this work bear the responsibility for publishing the MinT dataset and related code and data.

Data structure The structure of the provided MinT data is intentionally kept simple. All data is saved in CSV files or pandas DataFrames stored in pickle files. In Listing 1 we display how data for an individual sample can be loaded with minimal dependencies (*joblib* and *pandas*). We provide muscle activations in a range of $[0, 1]$, ground reaction forces and effective muscle forces. Data is provided with 50 fps, each dataframe is indexed by fractional timestamps. Columns are named meaningfully, the first 80 muscles belong to the lower body model, the following 322 muscels belong to the upper body model. The first and last 0.14 seconds are cut off since the muscle activation analysis is unstable towards the beginning and end of data. Since the data is generated in chunks of 1.4 seconds and muscle activation analysis can fail to succeed due to various factors, the provided data may contain gaps identified by missing data for certain time ranges.

```
>>> # First download and extract the dataset.
>>> # Example for sample
>>> #'BMLmovi/BMLmovi/Subject_11_F_MoSh/Subject_11_F_10_poses'
>>> import joblib
>>> joblib.load("muscle_activations.pkl")
      LU_addbrev_l    ...   TL_TR4_r  TL_TR5_r
0.14         0.016    ...      0.003     0.061
0.16         0.028    ...      0.005     0.070
0.18         0.033    ...      0.002     0.080
...            ...    ...        ...       ...
3.74         0.024    ...      0.020     0.028
3.76         0.016    ...      0.009     0.004
3.78         0.011    ...      0.003     0.000

[183 rows x 402 columns]

>>> joblib.load("grf.pkl")
      ground_force_right_vx  ...  ground_torque_left_z
0.14                 15.962  ...                   0.0
0.16                 10.596  ...                   0.0
0.18                  3.422  ...                   0.0
...                     ...  ...                   ...
3.72                 20.337  ...                   0.0
3.74                 21.572  ...                   0.0
3.76                 22.546  ...                   0.0

[182 rows x 18 columns]

>>> joblib.load("muscle_forces.pkl")
      LU_addbrev_l   ...  TL_TR4_r  TL_TR5_r
0.14         8.430   ...     0.153    11.652
0.16        15.345   ...     0.283    13.240
0.18        19.127   ...     0.143    15.240
...            ...   ...       ...       ...
3.72        14.437   ...     1.320     3.661
3.74        13.993   ...     1.270     5.330
3.76         9.346   ...     0.577     0.847

[182 rows x 402 columns]
```

Listing 1: Simplified loading of MinT samples with joblib and pandas.

The *musint* package To further facilitate the usage of the MinT dataset, we provide the *musint* package, a Python package that allows data to be loaded into a predefined torch dataset and allows simplified cross-referencing with BABEL dataset labels. Additionally, it includes functionality for sampling a sub-window of the data at a given framerate as well as identifying and handling any gaps in the data. A short example on the musint package usage is displayed in Listing 2.

The *musint* package can be installed via `pip install musint`. Additional insight can be found on the musint github page where we also provide a Jupyter notebook for displaying the data as well as additional information on muscle subsets:
https://github.com/simplexsigil/MusclesInTime

```
>>> # First download and extract the dataset.
>>> import os
>>> from musint.datasets.mint_dataset import MintDataset

>>> md = MintDataset(os.path.expandvars("$MINT_ROOT"))

>>> md.by_path("TotalCapture/TotalCapture/s1/acting2_poses")
MintData(path_id='s1/acting2', babel_sid=12906, dataset='
    TotalCapture', amass_dur=61.7, num_frames=1114, fps=50.0,
    analysed_dur=22.26, analysed_percentage=0.36, data_path='
    TotalCapture/TotalCapture/s1/acting2_poses', weight=72.1,
    height=169.2, subject='s1', sequence='acting2_poses',
    gender='male', has_gap=False, dtype=object))

>>> md.by_path("TotalCapture/TotalCapture/s1/acting2_poses").
    get_muscle_activations(time_window=(0.3,1.),
    target_frame_count=int(0.7*20.))
      LU_addbrev_l  ...  TL_TR4_r  TL_TR5_r
0.30         0.094  ...     0.000     0.020
0.36         0.094  ...     0.003     0.042
0.40         0.091  ...     0.000     0.027
 ...           ...  ...       ...       ...
0.90         0.093  ...     0.000     0.008
0.94         0.093  ...     0.000     0.000
1.00         0.094  ...     0.001     0.009

[14 rows x 402 columns]
```

Listing 2: Loading the MinT dataset with the python musint package.

A.2 Additional statistics and information

In Figure 9 we provide additional information on the data analyzed provided with Muscles in Time. Total Capture makes up a small part of the dataset with exceptionally long sequences. The Eyes Japan Dataset provides the largest contribution with 3.2h of analyzed recordings.

In Tables 3 and 4, we list larger muscle groups in the lower and upper body model as well as their function for human motion. Muscle groups or larger muscles can be represented by multiple simulated muscles, e.g. since such muscles are attached to multiple muscle locations or exert forces in varying directions. The *Gluteus Medius* muscle is an example with three simulated activations on each side of the body.

A.3 Design choices and more detailed limitations

The muscle-driven simulation, based on the approach by Falisse *et al*. [18], aims to ensure that muscle and skeletal dynamics align closely with given kinematic data while minimizing muscle effort. This process involves finding a solution within the problem space that satisfies an error tolerance and the number of collocation points, which depend on the dynamics of the kinematic data. Collocation points are used to discretize the continuous kinematic and dynamic equations into a finite set of points, making the optimization problem computationally feasible. To mitigate the risk of nonconvergent or nonmeaningful solutions, we implemented safeguards by restricting the deviation between the kinematic information before and after the optimization problem converges.

Given the computational complexity, we decided to use 50 collocation points per second and an error tolerance of 10^-3. On an Intel Xeon Gold 6230 with 96 GB RAM, processing 6 subsequences of 1.68 seconds (including 0.14 second buffers at start and end) in parallel took approximately a median time of 45 minutes. Figure 10 displays a distribution of sample-wise runtimes in a violin plot. Non-converging samples tend to have higher runtimes and can be found on the long tail on

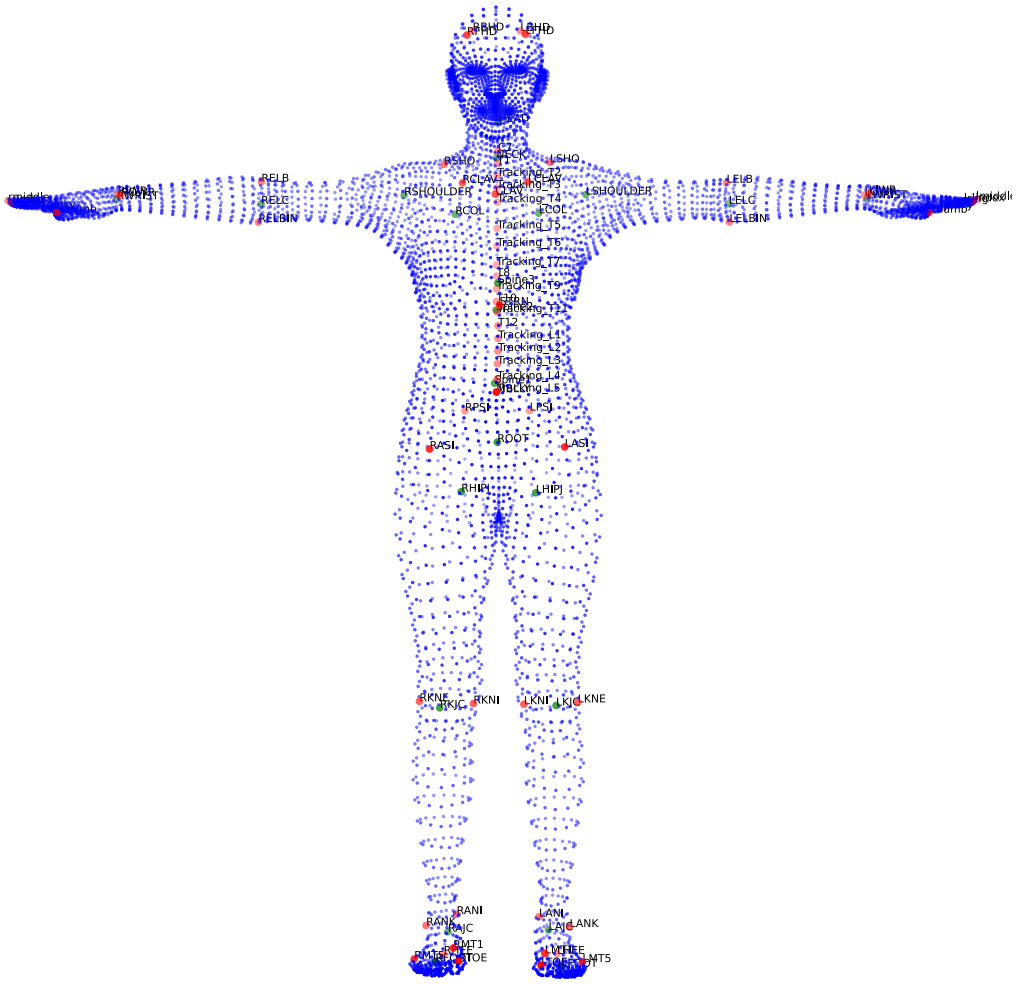

Figure 7: Virtual marker placement for transferring motions to OpenSim, enlarged from Figure 2.

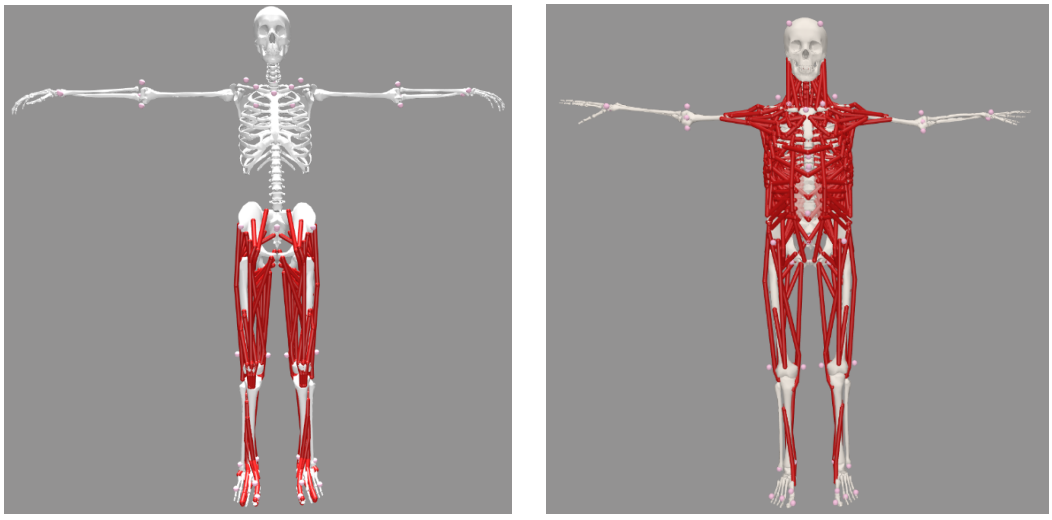

Figure 8: Lower body and upper body model, enlarged from Figure 2.

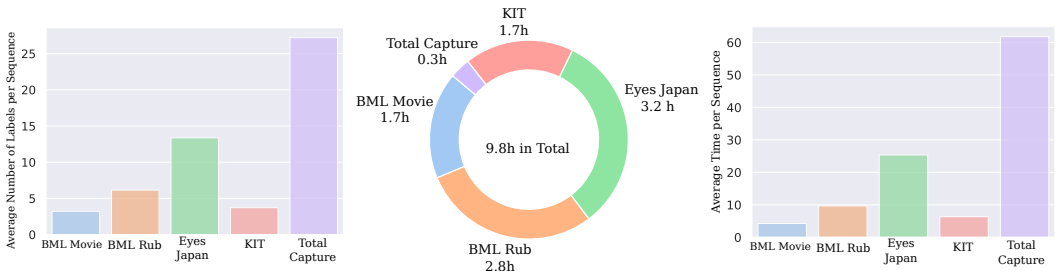

Figure 9: Average number of labels per sequence, composition of sub datasets and average sequence length.

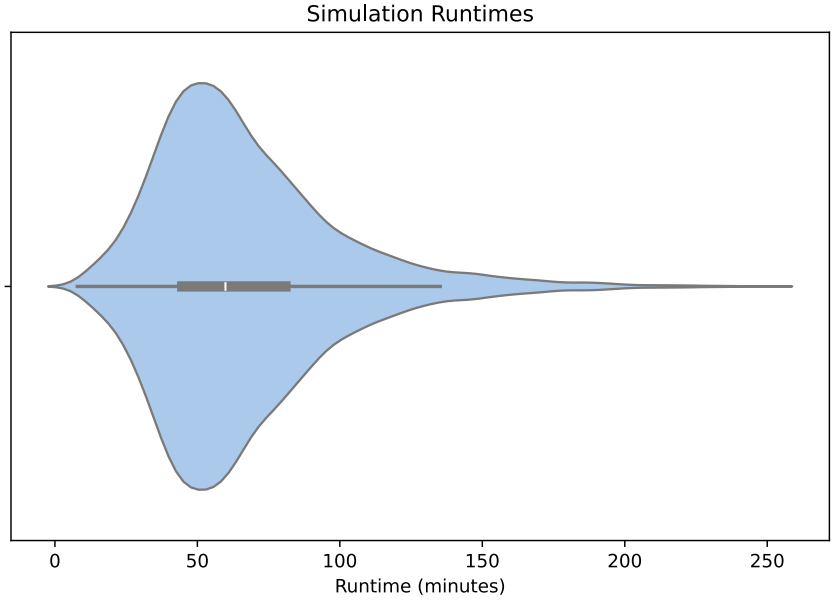

Figure 10: Analysis runtime distribution of the optimal trajectory problem described by Falisse *et al*. [18]. Subset of 10k runs.

the right. To manage the impact of unsuccessful simulations on the overall runtime, we limited the optimization problem to 2500 iterations and discard a sample if the optimization does not fall within error tolerance after this time. The AMASS sequences were divided into 1.4-second segments to mitigate a nonlinearly increasing runtime associated with longer motion sequences. After simulation, these segments were recombined into the original sequences, with muscle values smoothed at the connection points to ensure seamless transitions.

A challenge arose from minor variable distances between the AMASS body model and the ground, since the contact spheres provided by the OpenCap simulation are susceptible to changes in foot-ground distance. To provide similar foot-ground distances over all AMASS subjects, our pipeline automatically offsets the AMASS model depending on the lowest body marker over the time of the sequence.

Mapping AMASS motions to OpenSim models presented difficulties due to the numerous degrees of freedom in the Thoracolumbar model, complicating kinematic analysis. To safeguard the vertebral joints against aberrant movements, we constrained the range of motion for each vertebra, approximating the natural degrees of freedom in the vertebrae joints.

The MinT dataset was restricted to motions involving foot-ground contact only. Motions involving ground contact of other body parts or involving objects were excluded, except for motions that

Table 3: List of muscle groups modelled in the model by Lai et al. [33], which are analysed in the presented approach, and their functions [74].

Muscle	Function
Gluteus Maximus	Extension and rotation of the hip.
Gluteus Medius	Abduction and rotation of the thigh.
Gluteus Minimus	Abduction and rotation of the thigh.
Adductor Brevis	Adduction, flexion, and rotation of the thigh.
Adductor Longus	Adduction and flexion of the thigh.
Adductor Magnus	Adduction, flexion and rotation of the thigh.
Gracilis	Adduction, flexion and rotation of the thigh.
Semitendinosus	Flexion and rotation of the knee, as well as extension of the hip.
Semimembranosus	Flexion and rotation of the knee, as well as extension of the hip.
Tensor Fasciae Latae	Abduction and rotation of the thigh, as well stabilisation of the pelvis.
Piriformis	Rotation and extension of the thigh and abduction of thigh.
Sartorius	Flexion, abduction, and rotation of the hip and flexion of the knee.
Iliacus	Flexion of the hip.
Psoas	Flexion and rotation of the hip.
Rectus Femoris	Flexion of hip and extension of knee.
Biceps Femoris	Flexion of knee and extension of hip.
Medial Gastrocnemius	Flexion of foot and flexion of knee.
Lateral Gastrocnemius	Plantar flexion and knee flexion.
Tibialis Anterior	Dorsiflexion and inversion of the foot.
Vastus	Extension of the knee.
Extensor Digitorum Longus	Extension of toes and dorsiflexion of the foot.
Extensor Hallucis Longus	Extension of the big toe and dorsiflexion of the foot.
Flexor Digitorum Longus	Flexion of toes, as well as plantar flexion and inversion of the foot.
Flexor Hallucis Longus	Flexion of toes, as well as plantar flexion and inversion of the foot.
Peroneus (Fibularis)	Plantar flexion and eversion of the foot.
Soleus	Plantar flexion of the foot.

included throwing and lifting, which are particularly relevant for analyzing back muscle activation. In these cases, we assumed the objects' mass to be negligible, as the AMASS dataset does not provide this information.

A.4 Results for additional muscle subsets

To facilitate comparability to real world recordings as well as to other datasets, we define two muscle subsets of the lower body model, containing either 16 or eight of the most important lower body muscles for human locomotion. The subset `LAI_ARNOLD_LOWER_BODY_16` contains *left gluteus*

Table 4: List of muscle groups modelled in the model by Bruno et al. [3], which are analysed in the presented approach, and their functions [74].

Muscle	**Function**
Longissimus	Extension and rotation of the vertebrae.
Iliocostalis	Extension and flexion of the neck.
Semispinalis	Extension and rotation of the vertebrae.
Splenius	Extension and rotation of the vertebrae.
Sternocleidomastoid	Flexion and rotation of the head.
Scalenus	Elevation of ribs and flexion of the neck.
Longus Colli	Flexion of the neck and stabilisation of the cervical spine.
Levator Scapulae	Elevation and adduction of the scapula.
Quadratus Lumborum	Flexion the vertebral column.
Multifidus	Stabilisation cervical vertebrae.
Rectus Abdominis	Flexion of the lumbar spine.
External Oblique	Flexion and rotation of the trunk.
Internal Oblique	Flexion and rotation of the trunk.
Transversus Abdominus	Stabilisation of the trunk.

medius 1, *left adductor magnus ischial part*, *left gluteus maximus 2*, *left iliacus*, *left rectus femoris*, *left biceps femoris long head*, *left gastrocnemius medial head*, *left tibialis anterior*, *right gluteus medius 1*, *right adductor magnus ischial part*, *right gluteus maximus 2*, *right iliacus*, *right rectus femoris*, *right biceps femoris long head*, *right gastrocnemius medial head* and *right tibialis anterior* while the muscle subset `LAI_ARNOLD_LOWER_BODY_8` contains *left gluteus medius 1*, *left gluteus maximus 2*, *left rectus femoris*, *left biceps femoris long head*, *right gluteus medius 1*, *right gluteus maximus 2*, *right rectus femoris* and *right biceps femoris long head*. These subsets are also defined within the musint package.

In Table 5 we list the results of our 16 layer transformer model on these subsets.

Table 5: Human motion-to-muscle activation prediction results for the lower body model.

Motion	All muscles			Lower Body			Subset 16			Subset 8		
	RMSE↓	PCC↑	SMAPE↓	RMSE↓	PCC↑	SMAPE↓	RMSE↓	PCC↑	SMAPE↓	RMSE↓	PCC↑	SMAPE↓
overall	0.036	0.55	95.3	0.048	0.54	45.1	0.066	0.56	47.7	0.060	0.56	45.0
jump	0.052	0.64	100.7	0.051	0.71	52.3	0.059	0.71	55.5	0.056	0.70	54.2
kick	0.046	0.64	102.8	0.053	0.62	54.8	0.068	0.63	57.0	0.067	0.67	57.4
stand	0.033	0.56	97.5	0.046	0.58	45.0	0.062	0.61	47.5	0.052	0.59	43.6
walk	0.026	0.65	90.7	0.044	0.77	42.4	0.060	0.77	43.3	0.057	0.77	43.4
jog	0.033	0.71	99.0	0.046	0.71	51.1	0.063	0.75	51.8	0.062	0.71	52.7
dance	0.041	0.60	109.2	0.057	0.65	58.5	0.073	0.66	59.6	0.072	0.67	59.5

A.5 Training on Muscles in Action

We evaluate the generalizability of MinT by finetuning our 16-layer transformer architecture exclusively on the first and last transformer block and comparing the results with full training from scratch on Muscles in Action [9]. The motions in MIA were obtained with VIBE [32], a 3D pose estimation method performed on 2D images. The resulting motions are very noisy in contrast to the motions in AMASS which are the result of motion capture, inducing a significant domain gap. Table 6 shows our results. We find that limiting our training to the first and last transformer block results in very similar RMSE values compared to full fine-tuning, while PCC and SMAPE clearly displays a small

but significant advantage of the full fine-tuning strategy. Still, finetuning the first and last layer only affects some 8% of all trainable weights, and we see this as an indication for the transferability of the knowledge obtained by training on MinT.

Table 6: Human motion-to-muscle activation prediction results on Muscles in Action [9].

Motion	Full Fine-tuning			First and last layer		
	RMSE↓	PCC↑	SMAPE↓	RMSE↓	PCC↑	SMAPE↓
Overall	15.11	0.27	37.0	15.15	0.21	41.6
ElbowPunch	15.66	0.25	43.6	15.48	0.19	48.8
FrontKick	8.49	0.19	34.5	8.20	0.14	41.0
FrontPunch	8.47	0.38	29.8	8.22	0.36	36.3
HighKick	13.09	0.35	37.0	12.94	0.29	39.7
HookPunch	13.18	0.32	37.1	13.28	0.28	44.6
JumpingJack	13.79	0.27	28.5	13.42	0.23	29.5
KneeKick	12.32	0.25	37.3	12.26	0.16	43.0
LegBack	11.70	0.32	37.3	11.91	0.18	44.4
LegCross	13.89	0.17	42.7	13.84	0.11	48.9
RonddeJambe	15.81	0.20	39.5	15.50	0.17	42.6
Running	7.53	0.30	26.3	7.25	0.24	27.4
Shuffle	9.79	0.21	28.0	9.56	0.13	30.5
SideLunges	26.13	0.29	45.9	26.66	0.22	51.7
SlowSkater	20.15	0.26	42.1	20.81	0.19	47.2
Squat	22.68	0.26	44.9	22.76	0.21	48.2

A.6 Additional qualitative examples for MinT

Figure 6, in the main paper, lists two qualitative examples to display the muscle activation estimation quality of our best model. Additionally, Figures 11 to 17 show 48 randomly chosen samples from the test set.

A.7 Corrections

In line 266 and 267 we wrote

> In the construction of the dataset, some design choices had to increase simulation robustnees, [...]

while the correct text should be

> In the construction of the dataset, some design choices were made to increase simulation robustness, [...]

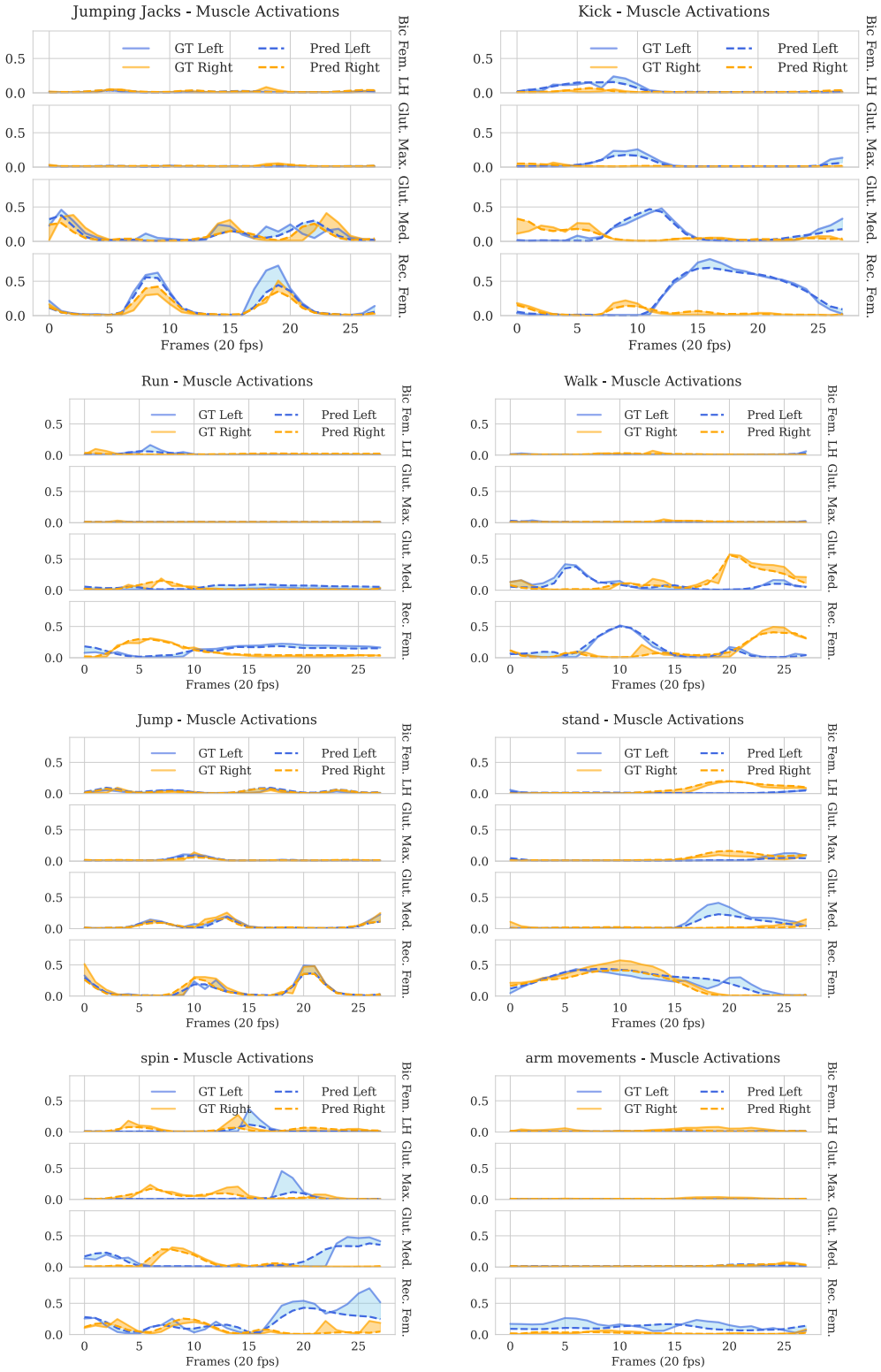

Figure 11: Muscle activation estimation with our 16 layer tranformer model.

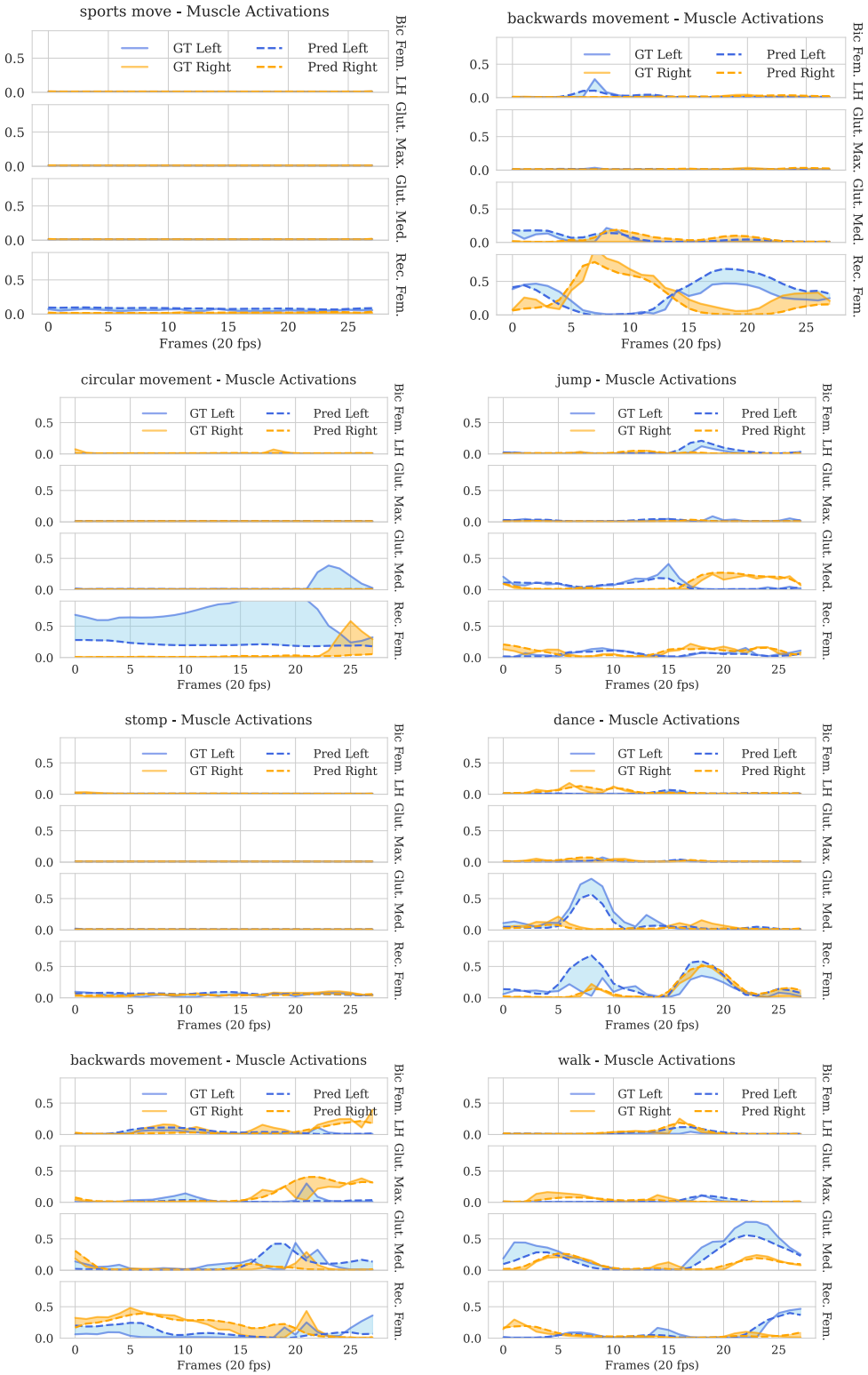

Figure 12: Muscle activation estimation with our 16 layer transformer model.

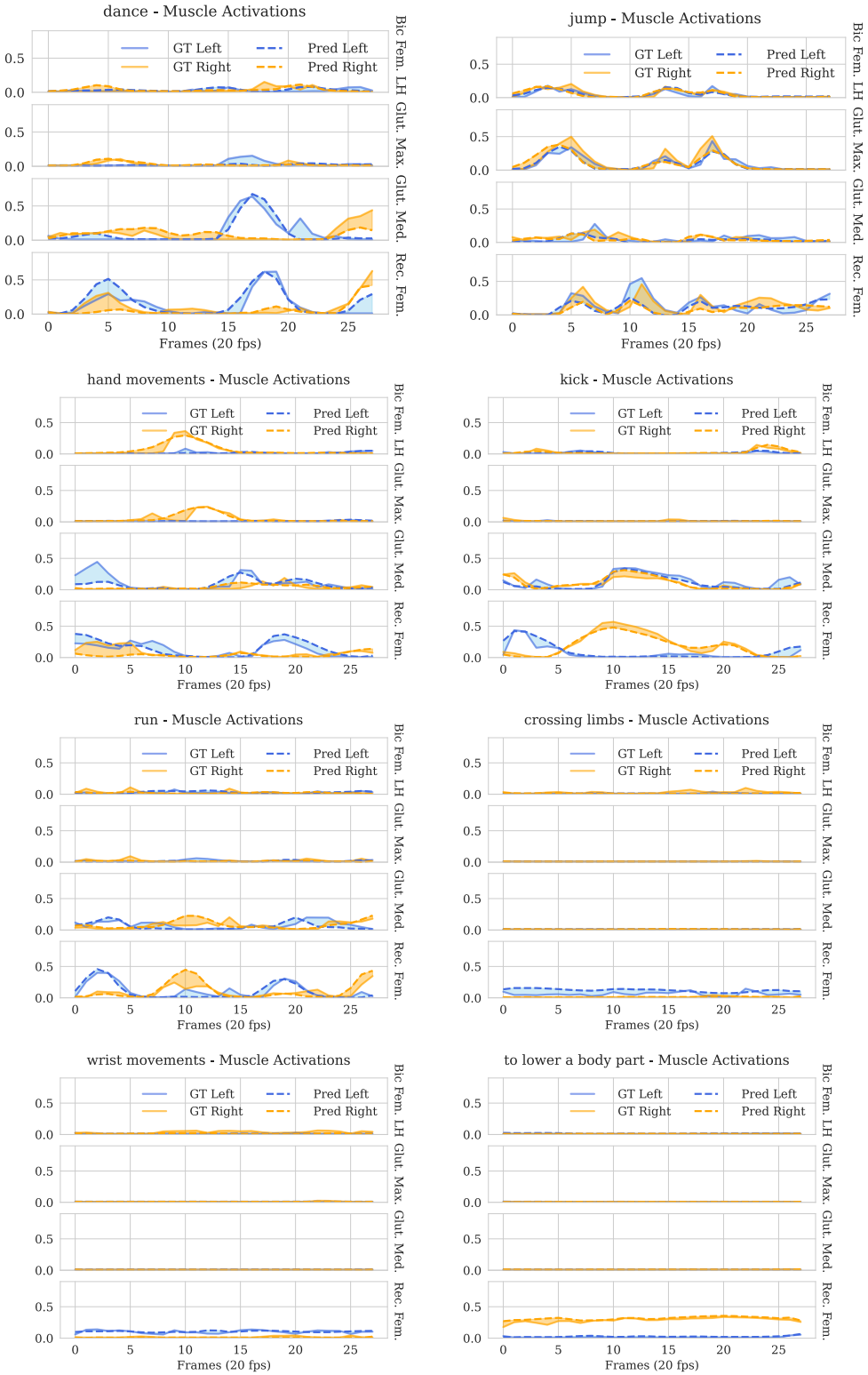

Figure 13: Muscle activation estimation with our 16 layer transformer model.

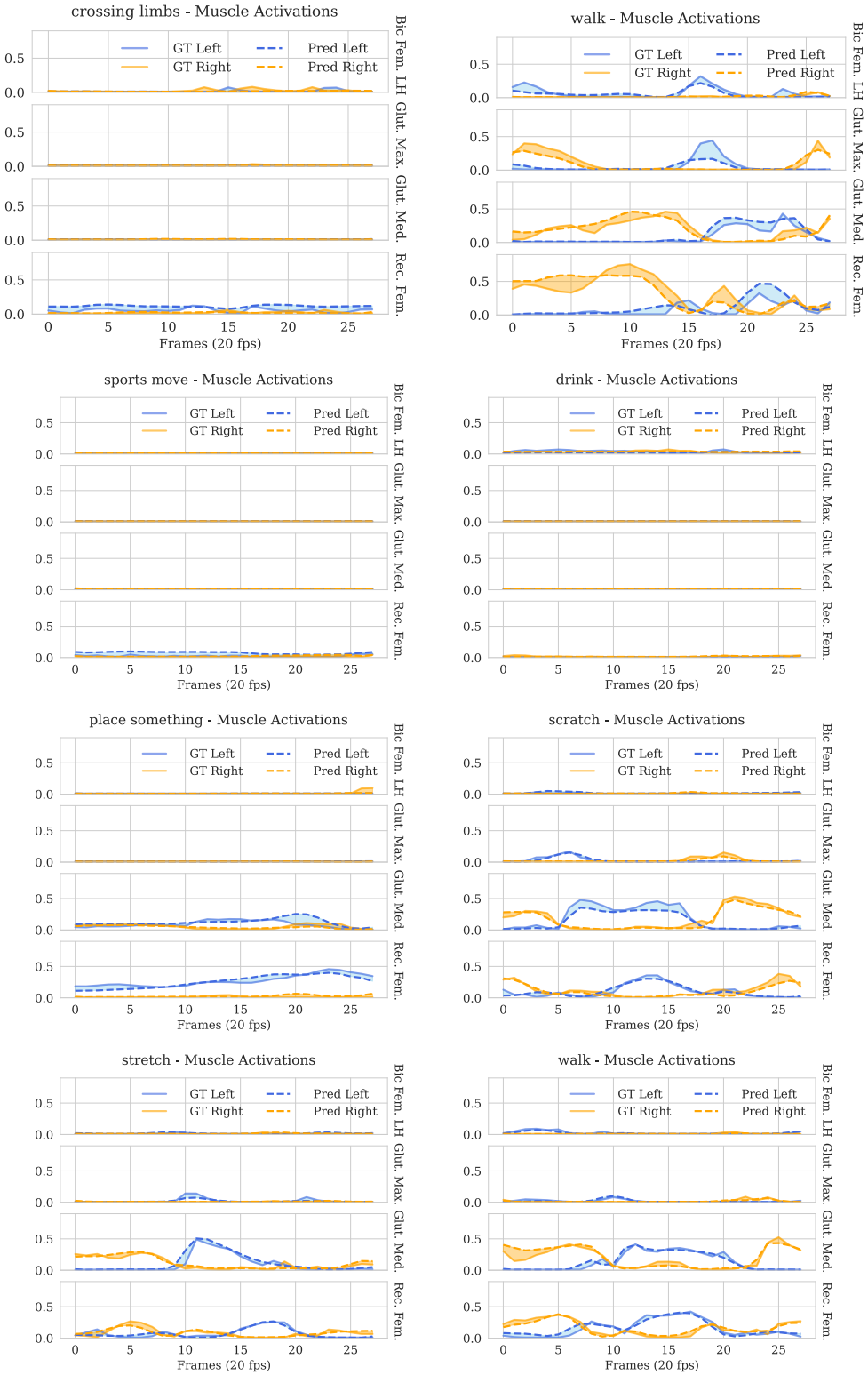

Figure 14: Muscle activation estimation with our 16 layer transformer model.

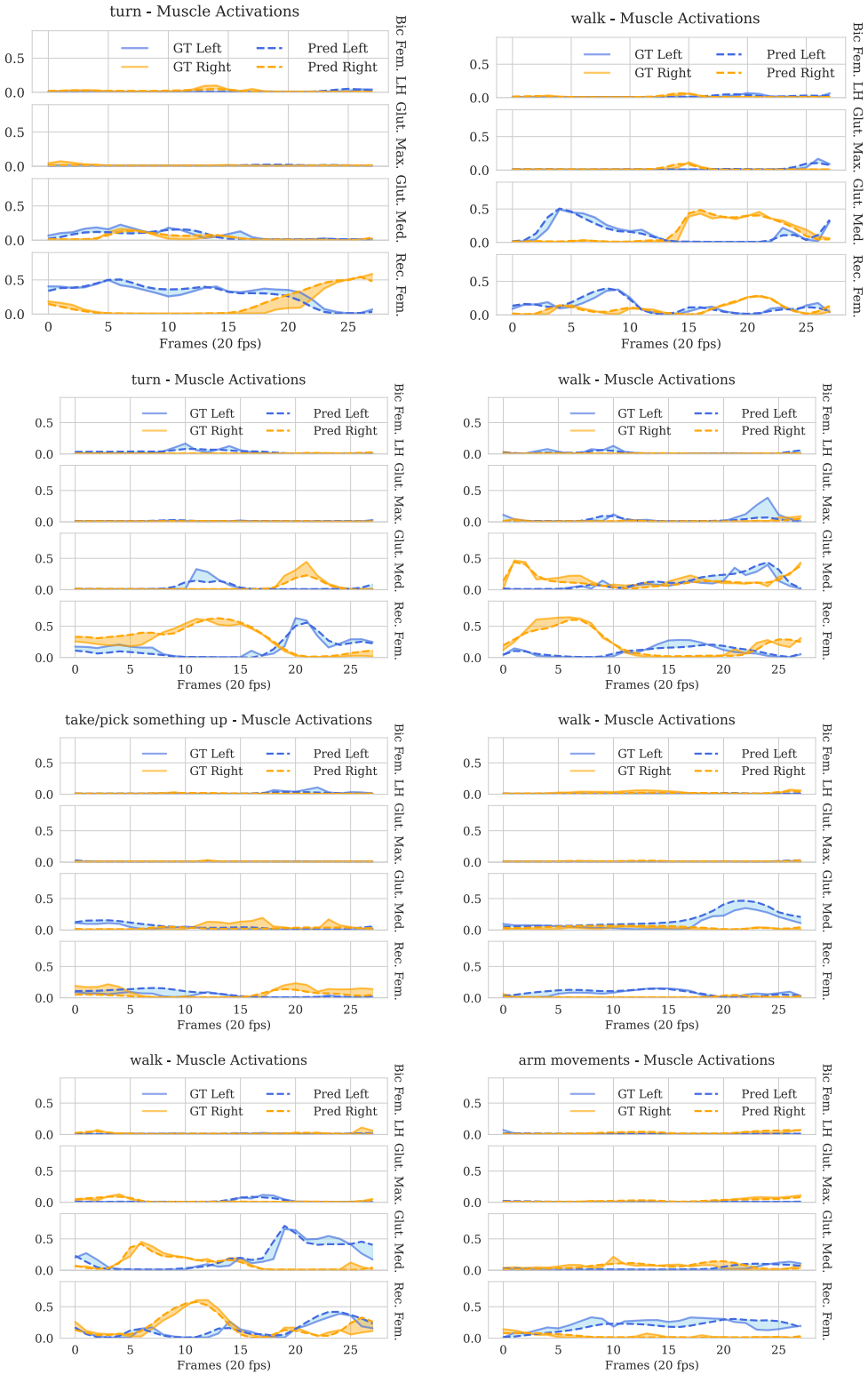

Figure 15: Muscle activation estimation with our 16 layer transformer model.

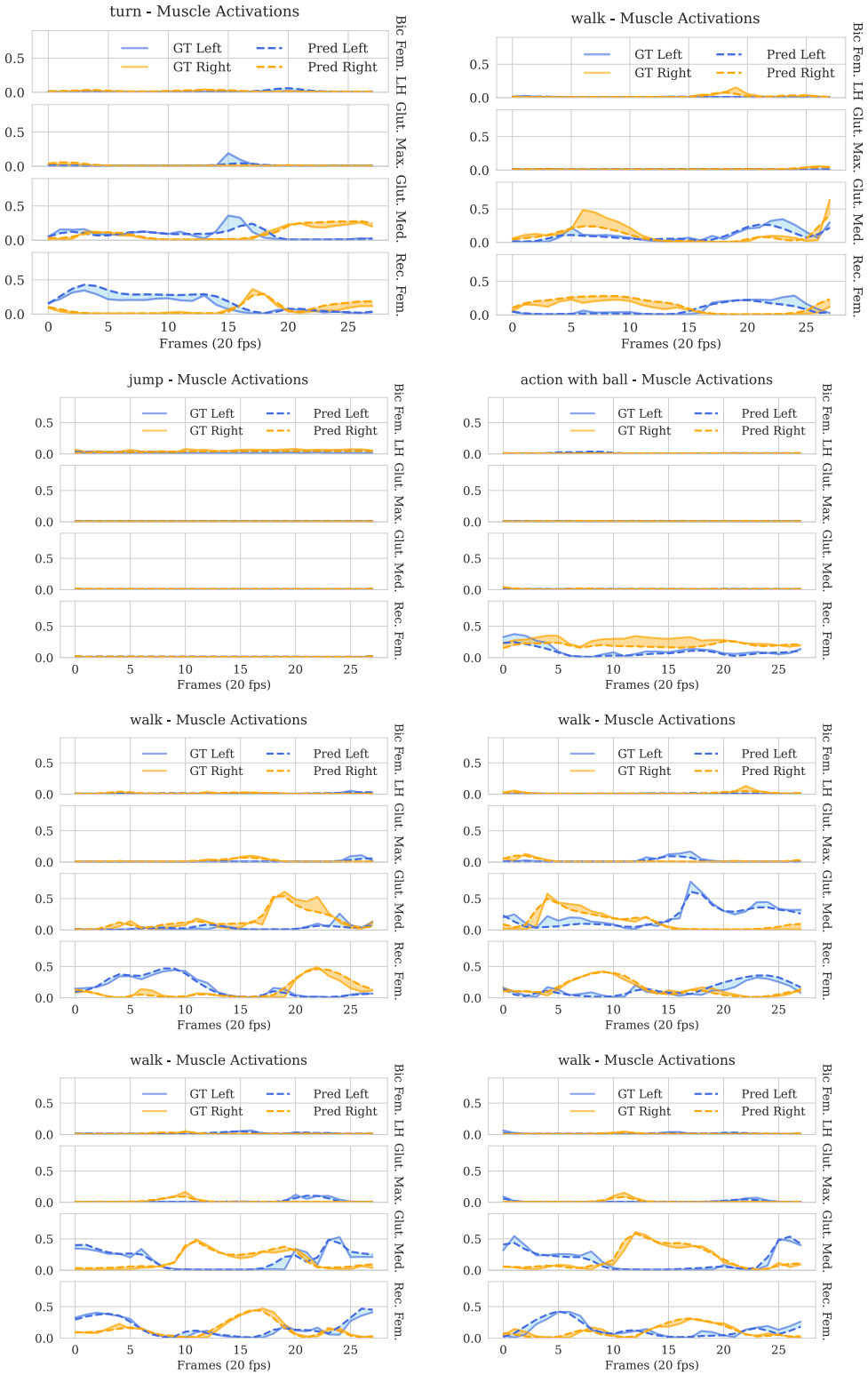

Figure 16: Muscle activation estimation with our 16 layer transformer model.

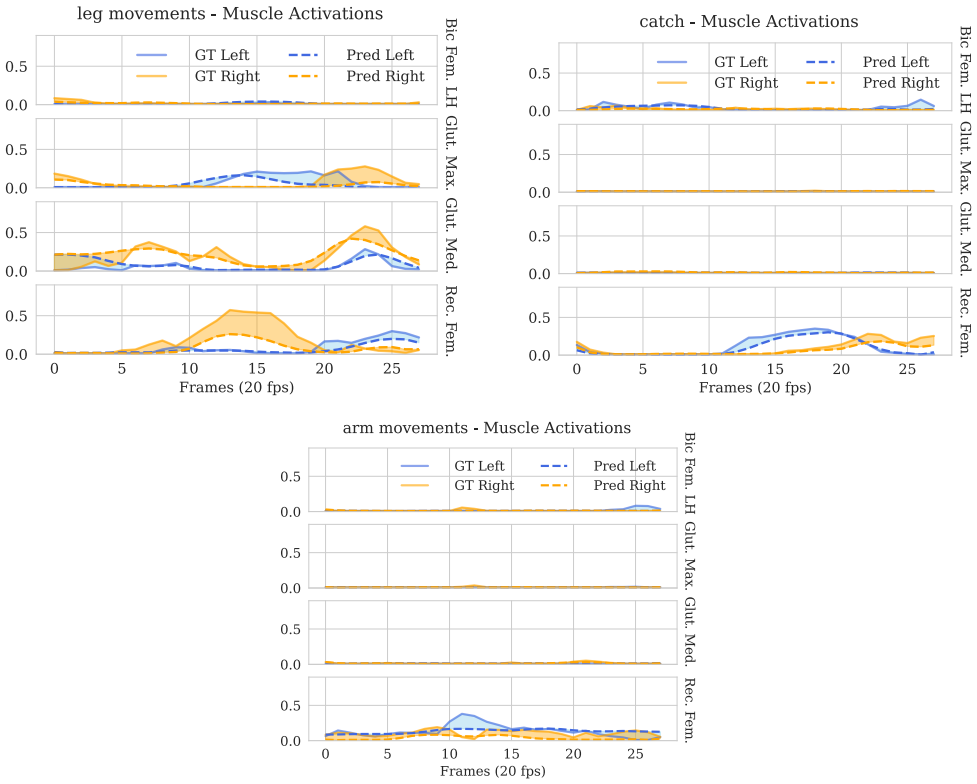

Figure 17: Muscle activation estimation with our 16 layer transformer model.