Proceedings of Machine Learning Research vol vvv:1–13, 2025

BIGE : Biomechanics-informed GenAI for Exercise Science

Shubh Maheshwari SHMAHESHWARI@UCSD.EDU
Anwesh Mohanty ANMOHANTY@UCSD.EDU
Yadi Cao YAC066@UCSD.EDU
Swithin Razu SRAZU@UCSD.EDU
Andrew McCulloch AMCCULLOCH@UCSD.EDU
Rose Yu ROSEYU@UCSD.EDU
University of California, San Diego
La Jolla, CA 92093

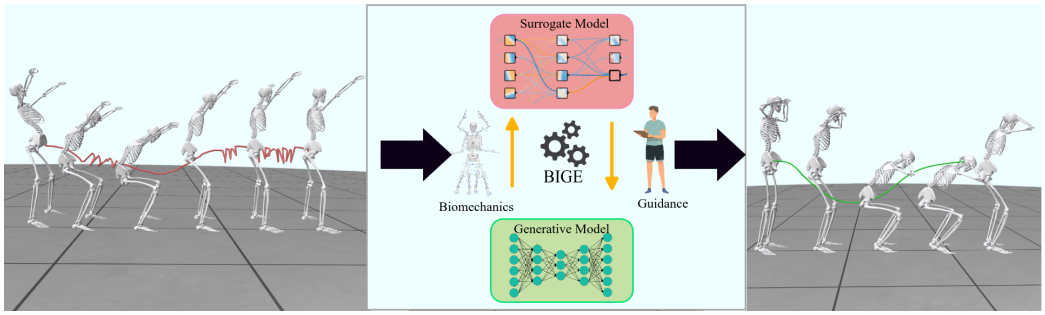

Figure 1: BIGE is a framework for generative models to adhere to clinician-defined constraints. To generate realistic motion, our method uses a biomechanically informed surrogate model to guide the generation process. Please refer to the project website: https://rose-stl-lab.github.io/UCSD-OpenCap-Fitness-Dataset/ for additional details.

Abstract

Proper movements enhance mobility, coordination, and muscle activation, which are crucial for performance, injury prevention, and overall fitness. However, traditional simulation tools rely on strong modeling assumptions, are difficult to set up and computationally expensive. On the other hand, generative AI approaches provide efficient alternatives to motion generation. But they often lack physiological relevance and do not incorporate biomechanical constraints, limiting their practical applications in sports and exercise science. To address these limitations, we propose a novel framework, BIGE, that combines bio-mechanically meaningful scoring metrics with generative modeling. BIGE integrates a differentiable surrogate model for muscle activation to reverse optimize the latent space of the generative model, enabling the retrieval of physiologically valid motions through targeted search. Through extensive experiments on squat exercise data, our framework demonstrates superior performance in generating diverse, physically plausible motions while maintaining high fidelity to clinician-defined objectives compared to existing approaches.

1. Introduction

Digital recovery and rehabilitation have become increasingly crucial in modern sports and exercise science. Healthcare professionals and researchers have explored various digital tools to simulate

physiologically proper motions that can guide practical training and rehabilitation processes. Physics-based simulations Keller et al. (2023); Lai et al. (2017); Rajagopal et al. (2016) have traditionally served as the primary tool to generate physiologically feasible motions by incorporating biomechanical constraints and principles. While these approaches ensure physical validity through proper integration of biomechanical models, they face significant limitations: the strong assumptions in biomechanical models can introduce bias in outputs, making them fail to capture the complexity in real-world data. Furthermore, they require substantial domain expertise to set up and tune, and the intense computational complexity makes them impractical for real-time clinical applications, such as generating guidance motions for patients on-the-fly.

Generative AI models have emerged as a promising alternative to traditional human motion simulation due to their ability to represent complex, real-world data. Advances in GenAI have enabled remarkable progress in human motion synthesis Zhang et al. (2023b,a); Tevet et al. (2022). However, these methods face limitations in clinical and rehabilitation settings. Naive generative models lack precise control over generated motions, while conditional models fail to guarantee physiological feasibility due to 1) the absence of meaningful biomechanical constraints in their conditioning and 2) reliance on motion representations, like SMPL Loper et al. (2015), that are not biomechanically grounded.

Motivated by these challenges, we propose a novel framework that guides the generative model with biomechanics metrics to generate physiologically meaningful human motions. Our framework builds on the idea of latent inceptionism Mordvintsev et al. (2015); Eckmann et al. (2022), which demonstrated score-guided generation in latent space for non-temporal tasks. We extend this concept to generate human motions that meet both observable biomechanical criteria, such as joint kinematics, and hidden criteria, such as muscle activations. By addressing both aspects, our approach ensures that generated motions minimize injury risk, making them suitable for rehabilitation and training.

Our methodology consists of three key components: 1) a generative model based on Vector-Quantized Variational Autoencoder (VQVAE) for learning diverse motion representations 2) a surrogate model to overcome the non-differentiable nature of inverse dynamics calculations for muscle activations and 3) a guidance policy that optimizes latent vectors based on physiologically meaningful objective scores. This combination enables our framework to generate high-fidelity, clinically relevant motions tailored to specific exercise goals.

Our main contributions are summarized below:

- We present a novel framework that combines human motion generative modeling with score-guided retrieval, capable of generating high-fidelity motions while accounting for biomechanical criteria, such as injury risk and personalized skeletal variations.
- We develop and validate crucial physiological scoring metrics for injury avoidance and rehabilitation. Our scores incorporate both observable and hidden biomechanical constraints, ensuring the clinical relevance of generated motions.
- Through extensive experiments, we demonstrate the superiority of our approach in generating squat motions that better satisfy the aforementioned objectives compared to existing state-of-the-art generative models.

2. Related Works

Biomechanical Simulations Traditional biomechanical simulations build upon foundational biomechanical modeling (such as Hill muscle model Hill (1938)) using musculotendon dynamics by Zajac (1989),. Several mature frameworks have emerged: OpenSim Seth et al. (2018); Delp et al. (2007), Moco Dembia et al. (2021) De Groote et al. (2016), a specialized model for upper body simulation Lee et al. (2009), OpenCap Uhlrich et al. (2022) for computing motions that best complement video input, and Nimble Werling et al. (2021) for differentiable simulation.

Human Motion Generation Machine learning has emerged as an alternative to traditional simulations. In particular, generative models for human motion synthesis Maheshwari et al. (2022); Gupta et al. (2023); Guo et al. (2020); Petrovich et al. (2021); Tevet et al. (2022). Recent works have explored approaches to control the motion generation process using text Zhang et al. (2023a); Tevet et al. (2022); Kalakonda et al. (2022). Although these methods have significantly improved motion synthesis quality, they generate physiologically unexplainable motions because their representation (SMPL Loper et al. (2015) and its variants Guo et al. (2022)) inherently lacks biomechanical understanding. However, none of these works incorporate biomechanically meaningful constraints, which could be beneficial for clinical applications such as rehabilitation.

Score-Guided Retrieval Generation Although not in the field of human motion generation, LIMO Eckmann et al. (2022) shows great promise in retrieving generated targets that satisfy certain objective constraints. Although LIMO demonstrated success in molecular properties, this scenario contains only static generation (compound formulation) and constraints (chemical properties). It has not been studied or validated in scenarios involving highly dynamic systems, such as human motion. Motivated by LIMO, we adopt a similar approach and design several crucial criteria scores for injury avoidance and rehabilitation. To our knowledge, the only similar work is Yao et al. (2022). However, this work only considers a hidden criterion: muscle activation. In contrast, our diverse scoring criteria cover both observable and hidden states.

3. Methodology

Figure 1 provides a high-level overview of BIGE, our biomechanic-informed generative model. BIGE performs score-guided optimization to refine the latent space representations of the generative model subject to biomechanical constraints to generate high-quality motion samples. Section 3.1 details our score-guided generative model, which leverages property-based objectives to refine motion generation. Section 3.2 introduces the surrogate model, which incorporates biomechanical data to predict physiologically relevant features such as muscle activations. Finally, we discuss how biomechanical knowledge is used to model properties and guide the generative process toward producing clinically meaningful exercise motions.

3.1. Score-Guided Generative Model

Our score-guided generative model combines property-based guidance to optimize latent space representations to generate motion sequences. The motion sequence represents physiologically meaningful joint kinematics $K \in \mathbb{R}^{T \times D}$, where T is the number of time steps and D is the degree of freedom of the biomechanical mode.

Training: We train a VQVAE van den Oord et al. (2017) to learn a compact latent representation of the joint kinematics. The VQVAE consists of an encoder f_{enc}, a decoder f_{dec}, and a discrete codebook $C = \{c_i\}_{i=1}^{N}$ where each $c_i \in \mathbb{R}^{d_c}$, and d_c is the dimensionality of the codebook and N is number of codebook vectors. The encoder maps K to latent representations $\mathbf{Z}$ which are quantized before passing through the decoder. For quantization, each latent feature z_i in Z is mapped to the nearest codebook vector $c_j \in C$ using $\hat{Z}_i = \arg\min_{c_j \in C} ||Z_i - c_j||_2$, resulting in quantized latent features $\hat{Z} = [\hat{Z}_1, \hat{Z}_2, \dots]$. The reconstructed motion $\hat{K}$ is obtained by passing $\hat{Z}$ through the decoder, $\hat{K} = f_{\text{dec}}(\hat{Z})$. The VQVAE training objective $\mathcal{L}_{vq}$ combines a reconstruction loss, a temporal smoothing term, and a commitment loss:

$$\begin{aligned} &\mathcal{L}_{vq} = \mathcal{L}_{\text{rec}} + \lambda_{\text{temp}}\mathcal{L}_{\text{temp}} + \lambda_{\text{commit}}\mathcal{L}_{\text{commit}}, \\ &\mathcal{L}_{\text{rec}} = ||K - \hat{K}||_2, \quad \mathcal{L}_{\text{temp}} = \sum_{t=2}^{T} ||\hat{K}_t - \hat{K}_{t-1}||_2, \quad \mathcal{L}_{\text{commit}} = ||Z - \text{sg}[\hat{Z}]||_2, \end{aligned} \tag{1}$$

where sg is the stop-gradient operator. To prevent codebook collapse, the codebook is updated using an exponential moving average (EMA) as follows: $C^{\text{iter}} \leftarrow \lambda_{\text{EMA}} C^{\text{iter-1}} + (1 - \lambda_{\text{EMA}}) C^{\text{new}}$, where λ_{EMA} is the EMA constant, and inactive codes are reassigned based on input data.

Guidance: Figure 2 provides a high level overview of our guidance framework. To constrain the generated motion to exhibit desired properties, we draw inspiration from "Latent Inceptionism" Eckmann et al. (2022). We use the property predictor, g_α (details in Sec.3.2) to guide latent vectors toward favorable motions. The weights of the decoder f_{dec} of the model and property predictor g_α are frozen, and the optimization process focuses on finding optimal latent variables $\mathbf{Z}$ that generate motions satisfying clinician-defined constraints. Latent vectors Z are initially sampled from a normal distribution $N(0, I)$, and optimized using the Adam optimizer to minimize a regression loss:

$$L_{\text{LIMO}} = ||\text{GT} - g_\alpha(f_{\text{dec}}(Z))||_2, \tag{2}$$

where GT represents the ground truth obtained from the reference dataset, and $g_\alpha(f_{\text{dec}}(z))$ is the predicted property. To ensure that the optimized Z remains close to the training distribution, we introduce a proximity loss. This term minimizes the L2 distance between the optimized Z and the nearest latent vector from the training data: $L_{\text{proximity}} = ||Z - z_{\text{nearest}}||_2$, where z_{nearest} is the closest latent vector from the training dataset. The total loss function combines these terms:

$$L_{\text{total}} = L_{\text{LIMO}} + \beta L_{\text{proximity}} \tag{3}$$

where β is a tunable hyperparameter. This formulation ensures that the guidance policy generates motions with the desired properties while maintaining proximity to the training data distribution.

3.2. Surrogate Model for Muscle Activation

Muscle activations reveal the relationships between muscle coordination, kinematics, and clinically relevant metrics, enabling the discovery of optimal strategies that may be difficult for humans to identify unaided. We leverage muscle-driven simulations (Falisse et al. (2019)) and data from Uhlrich et al. (2022) to generate ground truth for training models and designing scoring functions. To enable physiologically relevant feedback for motion generation, we propose a surrogate model that efficiently estimates muscle activations from joint kinematics. This model eliminates the

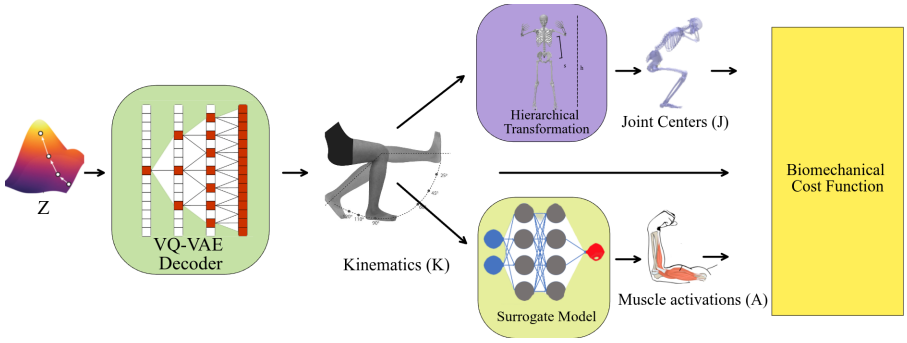

Figure 2: Score Calculation: Latent variables **Z** (sampled randomly) are decoded using the decoder to generate joint kinematics output K. Next, hierarchical transformations are applied to the biomechanical model, to compute joint centers **J**. Then a surrogate model predicts muscle activations $\hat{A}$. Finally, clinician-defined constraints are imposed on the derived variables, generating a score S using the property predictor g_α

computational burden of traditional simulations, making it feasible for VQVAE training while ensuring anatomical plausibility and biomechanical consistency. Additionally, the surrogate model bypasses the need for complex gradient approximations typically required for non-differentiable physical simulations.

The surrogate model is a sequence-to-sequence mapping, $f_{\text{sur}} : \mathbb{R}^{T\times D} \to \mathbb{R}^{T\times M}$, where $\hat{K} \in \mathbb{R}^{T\times D}$ represents the generated kinematics from our generative model, and $\mathbf{y} \in \mathbb{R}^{T\times M}$ corresponds to predicted muscle activations. A 3-layer transformer architecture with multi-head attention captures both local and long-range temporal dependencies, ensuring accurate predictions. Sigmoid activation functions on the output ensure the muscle activations are bounded within the physiologically valid range of $[0, 1]$. To train the surrogate model, we minimize the L1 loss between predicted and true muscle activations:

$$\mathcal{L} = \frac{1}{TM} \sum_{t=1}^{T} \sum_{m=1}^{M} \left| A_{t,m} - \hat{A}_{t,m} \right|,$$

where $A_{t,m}$ and $\hat{A}_{t,m}$ denote the simulation and predicted activations, respectively, for muscle m at time step t. Further training details, including data preprocessing, optimization parameters, and hyperparameter settings, are provided in the supplementary material.

3.3. Biomechanical Constraints and Guidance Policy

Our framework allows clinicians to impose multiple biomechanical constraints and criteria on the generated motion, ensuring adherence to physiologically meaningful properties. These constraints can target various aspects of the motion, including joint kinematics (pelvis tilt, angular velocity ω), joint center dynamics (center of mass (COM) velocity and acceleration), and muscle activations. For the purpose of this paper, we will be focusing on the 'SQUAT' motion. This framework can be extended to other motions as well since other motions will also use constraints relating to certain joint angles or muscle activations. To incorporate these criteria, we define a user-defined objective function, $g_\alpha : K \to S$, where S is a scalar score indicating how well the generated motion satisfies the desired properties. The final objective is formulated as a weighted sum of individual criteria.

Joint Kinematics: Joint kinematics offer a direct means to enforce meaningful constraints on properties like pelvic tilt, asymmetry and angular velocity (ω). Pelvic tilt encourages peak muscle

activation by maximizing pelvic tilt at specific moments during the motion. Asymmetry minimizes the difference in flexion angles of corresponding left and right joints (hip, knee, and ankle) at all time steps to enforce symmetrical motion. ω reduces temporal jitter by applying a temporal smoothing constraint that minimizes joint angular velocity and prevents the generation of unnatural motion.

Joint Centers: To ensure smooth and natural motion, we impose constraints on the dynamics (velocity and acceleration) of COM of joint centers. This reduces jitter by minimizing velocity and acceleration variations in the COM trajectory. Acceleration is computed using the Laplacian operator applied to the COM trajectory.

Muscle Activation: To promote physiologically relevant muscle activations, we enforce range constraints on the 'Vastus Medialis' muscles. This leads to deeper squats by restricting the activations of the left and right vastus medialis muscles within a user-defined range.

By combining these criteria, our framework ensures that the generated motions adhere to the specified biomechanical properties, enabling the generation of clinically meaningful and physiologically consistent motion sequences. The definitions and implementations for the above biomechanical constraints are summarized in Table.1. The final property predictor can be defined as:

$$g_\theta = \alpha_{\text{tilt}}\mathcal{L}_{\text{tilt}} + \alpha_{\text{asymm}}\mathcal{L}_{\text{asymm}} + \alpha_\omega\mathcal{L}_\omega + \alpha_{v_{\text{COM}}}\mathcal{L}_{v_{\text{COM}}} + \alpha_{a_{\text{COM}}}\mathcal{L}_{a_{\text{COM}}} + \alpha_{\text{slide}}\mathcal{L}_{\text{slide}} + \alpha_{\text{MAC}}\mathcal{L}_{\text{MAC}} \quad (4)$$

Metric	Definition	Guidance Loss
Pelvis Tilt (degs)	$\max(K_{\text{pelvis}})$, where K_{pelvis} is the pelvis tilt angle at peak activation	$\mathcal{L}_{\text{tilt}} = \frac{1}{T}\sum_{t=1}^{T}\hat{K}_{t,\text{peak\_timestep,pelvis\_tilt\_index}}$
Asymmetry (degs)	$\text{mean}(\sum_{(j_l,j_r)\in\text{leg joints}}\lvert K_{j,\text{left}} - K_{j,\text{right}}\rvert)$, comparing the angles of left and right leg joint	$\mathcal{L}_{\text{asymm}} = \frac{1}{T}\sum_{t=1}^{T}\sum_{j_l,j_r\in\text{leg joints}}(\hat{K}_{\text{left}} - \hat{K}_{\text{right}})^2$
Angular vel. (degs/s)	$\text{mean}(\sum_{\theta\in K^{\text{angles}}}\frac{\text{d}\theta}{\text{d}t})$, where K^{angle} is joint angles in Kinematics	$\mathcal{L}_\omega = \frac{1}{T}\sum_{t=1}^{T}\sum_{\theta\in K^{\text{angles}}}(\hat{K}_{t,\theta} - \hat{K}_{t,\theta})^2$
COM vel. (m/s)	$\text{mean}(\frac{\lvert\lvert\text{d}K^{trans}\rvert\rvert}{\text{d}t})$, where $K^{trans} \in \mathbb{R}^3$ pelvis translation relative to the ground	$\mathcal{L}_{v_{\text{COM}}} = \frac{1}{T-1}\sum_{t=1}^{T-1}(\hat{K}_{t+1}^{trans} - \hat{K}_t^{trans})^2$ is the
COM acc. (m/s^2)	$\text{mean}(\frac{\lvert\lvert\text{d}^2K^{trans}\rvert\rvert}{\text{d}t^2})$, where ∇^2 is the Laplacian operator	$\mathcal{L}_{a_{\text{COM}}} = \frac{1}{T-2}\sum_{t=2}^{T-1}(\nabla^2\hat{K}^{trans})^2$
Foot Sliding	$\text{mean}(\lvert\lvert\frac{\text{d}p_{\text{lowest}}}{\text{d}t}\rvert\rvert)$, ensures the lowest point p_{lowest} does not slide on the ground	$\mathcal{L}_{\text{slide}} = \frac{1}{T-1}\sum_{t=1}^{T-1}(\mathbf{p}_{\text{lowest},t+1} - \mathbf{p}_{\text{lowest},t})^2$
Muscle activation constrain (MAC)	Ensure vastus medialis muscle activation $\hat{A}_{\text{vas\_med}}$ remains in the given range [low,high]	$\mathcal{L}_{\text{MAC}} = \frac{1}{T}\sum_{t=1}^{T}\max(x - \text{high}, \text{low} - x, 0)$

Table 1: Summary of loss constraints and their implementations.

To compute joint centers from kinematics, we employ Werling et al. (2021) to perform hierarchical transformations. At each time step, hierarchical transformations deform the biomechanical model to the desired pose using transformation matrices for the kinematic chain. For joint j, transformation matrix $\mathbf{T}_j = \mathbf{R}_j\mathbf{T}_{j-1} + \mathbf{t}_j$, where $\mathbf{R}_j$ is the rotation matrix derived from joint angles, $\mathbf{T}_{j-1}$ is the parent joint's transformation matrix, and $\mathbf{t}_j$ is the translation vector for joint displacements. Joint centers $\mathbf{J}_j$ for each joint j are extracted from $\mathbf{T}_j$ as $\mathbf{J}_j = \mathbf{T}_j\left[0,0,0,1\right]^T$

4. Experiments Results

4.1. Data collection

We collect data from 80 subjects performing squats and use OpenCap (Uhlrich et al. (2022)) to perform muscle driven simulations. We use the Lai et al. (2017) musculoskeletal which includes 80 muscles for lower-limb coordinates. Ground reaction forces are modeled through six foot-ground contact spheres. The tracking simulation is formulated as optimal control problems that aim to

identify muscle excitations that minimize a cost function subject to constraints describing muscle and skeletal dynamics. The cost function includes terms for muscle activations, torque motors and tracking terms for joint positions, velocities and accelerations. All the settings are kept for 'squat' except tolerance 10^{-3}. 60 subjects are used for training and 20 for validation.

4.2. Baselines

To evaluate the effectiveness of our guidance pipeline, we compare BIGE against traditional motion generation methods as well as state-of-the-art human motion generation AI models:

- **Reference**: Data collected from muscle-driven simulation.
- **Markerless Motion Capture (MoCap)** OpenCap leverages pose estimation algorithms to identify body landmarks from videos to estimate kinematics. These kinematics represent the motion data obtained from a markerless motion capture system using predictive models as opposed to generative models. Note that this data does not contain muscle activation.
- **MDM** Tevet et al. (2022): A classifier-free, diffusion-based generative model for the human motion domain, conditioned on language inputs and implemented using transformers.
- **T2M-GPT** Zhang et al. (2023a): A language-guided generative framework that combines a Vector Quantized-Variational AutoEncoder (VQ-VAE) with a Generative Pre-trained Transformer (GPT) to generate human motion from textual descriptions.
- **VQ-VAE** van den Oord et al. (2017): A VQ-VAE trained directly on the joint kinematics of a biomechanical model.

For MDM and T2M-GPT, we generate high-quality squat motions by providing the models with the textual prompt: "a person is doing squats with hands behind head". Since these models output data in SMPL format, we apply inverse kinematics (IK) using Werling et al. (2021) to convert the generated motions into joint kinematics format. For VQ-VAE, we directly sample latent vectors from the encoder's latent space to produce squat motions, which are inherently in joint kinematics format.

4.3. Evaluation Metrics

We employ a set of metrics designed to capture both realism and biomechanical accuracy. Realism is assessed using fidelity and diversity metrics, which measure how closely the generated motions align with real-world data while maintaining natural variability. Floor interaction metrics, including penetration, floating, and sliding, evaluate the physical plausibility of the generated motions by analyzing their interaction with the ground. These metrics are critical for ensuring that the motions are not only visually realistic but also biomechanically consistent and suitable for clinical applications.

Fidelity: To evaluate the fidelity of the generated motion data, we compute the 2-Wasserstein distance, between the real data distribution, P and the generated data distribution, Q. This metric compares the aggregated distributions by considering their means, $\mu_{\text{real}} = \mathbb{E}x \sim P[x]$ and $\mu_{\text{gen}} = \mathbb{E}x \sim Q[x]$, as well as their variances $\sigma^2_{\text{real}} = \mathbb{E}x \sim P[(x - \mu_{\text{real}})^2]$ and $\sigma^2_{\text{gen}} = \mathbb{E}x \sim Q[(x - \mu_{\text{gen}})^2]$. Specifically, the Wasserstein distance is calculated as:

$$[W_2^2(P, Q) = (\mu_{\text{real}} - \mu_{\text{gen}})^2 + (\sigma_{\text{real}} - \sigma_{\text{gen}})^2] \tag{5}$$

Model	Realism		Floor metrics		
	Fidelity ↓	Diversity →	Penetration (cm) ↓	Floating (cm) ↓	Sliding ($\frac{m}{sec}$) ↓
Reference	-	1.736	$0.003^{\pm 0.057}$	$0.000^{\pm 0.021}$	$0.035^{\pm 0.024}$
MoCap	$\underline{7.855}^{\pm 9.902}$	$\underline{1.773}^{\pm 0.015}$	$\mathbf{0.501}^{\pm 0.987}$	$\mathbf{0.759}^{\pm 1.649}$	$\mathbf{0.209}^{\pm 0.161}$
MDM	$257.95^{\pm 105.92}$	$1.26^{\pm 0.05}$	$2.685^{\pm 3.840}$	$3.216^{\pm 4.636}$	$1.395^{\pm 2.672}$
T2M-GPT	$459.49^{\pm 257.30}$	$1.45^{\pm 0.05}$	$\underline{1.997}^{\pm 2.938}$	$2.440^{\pm 3.606}$	$3.770^{\pm 4.758}$
VQVAE	$45.99^{\pm 4.67}$	$1.82^{\pm 0.03}$	$2.035^{\pm 3.016}$	$\underline{2.331}^{\pm 3.506}$	$1.351^{\pm 0.830}$
BIGE	$\mathbf{6.48}^{\pm 4.46}$	$\mathbf{1.76}^{\pm 0.03}$	$2.327^{\pm 3.659}$	$2.379^{\pm 3.561}$	$\underline{0.291}^{\pm 0.255}$

Table 2: Performance of different models based on realism, and floor metrics. The right arrow (→) indicates that the values should be close to the Reference data. Down arrow (↓) means a value closer to 0 is better.

Model	Pelvis tilt(degs) →	Asymmetry(degs) →	ω (degs/s) →	COM vel.(m/s)→	COM acc. (m/s^2) →	Constrain (0-1)↓
Reference	-23.40	2.42	90.04	0.54	2.65	0.08
MoCap	$-28.13^{\pm 6.03}$	$\underline{2.76}^{\pm 0.59}$	$\mathbf{96.24}^{\pm \mathbf{13.96}}$	$\underline{0.67}^{\pm 0.14}$	$\underline{30.45}^{\pm 5.55}$	$0.03^{\pm 0.03}$
MDM	$0.42^{\pm 1.87}$	$\mathbf{2.25}^{\pm \mathbf{0.28}}$	$554.10^{\pm 112.77}$	$3.39^{\pm 0.79}$	$335.86^{\pm 81.37}$	$\underline{0.03}^{\pm 0.02}$
T2M-GPT	$4.16^{\pm 2.42}$	$4.73^{\pm 0.29}$	$918.14^{\pm 42.06}$	$6.09^{\pm 0.48}$	$597.95^{\pm 54.60}$	$\mathbf{0.02}^{\pm \mathbf{0.01}}$
VQVAE	$\underline{-20.69}^{\pm 0.51}$	$1.12^{\pm 0.03}$	$37.78^{\pm 0.42}$	$1.59^{\pm 0.08}$	$128.24^{\pm 6.49}$	$0.16^{\pm 0.01}$
BIGE	$\mathbf{-23.09}^{\pm \mathbf{1.54}}$	$2.09^{\pm 0.14}$	$\underline{56.95}^{\pm 1.94}$	$\mathbf{0.51}^{\pm \mathbf{0.04}}$	$\mathbf{27.32}^{\pm \mathbf{3.88}}$	$0.06^{\pm 0.01}$

Table 3: Performance evaluation on guidance metrics. The right arrow (→) indicates that the values should be close to the Reference data. Down arrow (↓) means a value closer to 0 is better.

Diversity: To quantify the diversity of the generated motion data, we compute the Shannon entropy $H(X) = -\sum_{i=1}^{N} p_i \log p_i$ of both the real and generated datasets, where N is the number of bins in the histogram and p_i is the probability associated with bin i. Entropy measures the uncertainty or variability in the data distribution; higher entropy indicates greater diversity in the generated samples. The diversity metric D is defined as the absolute difference between the entropies of the real and generated datasets, ensuring that the generative model produces outputs with a comparable level of variability as observed in the real dataset.

Floor metrics To evaluate the physical plausibility of generated motions, we also report floor based metrics proposed in Yuan et al. (2023). The metrics are calculated using the joint centers. The ground is estimated using the median of the minimum y-coordinate of all time steps for all samples.

- **Penetration**: Displacement below the ground at each time step for all subjects defined as: $\text{Penetration}(\overline{P}) = \frac{1}{T}\sum_{t=1}^{T} \max(0, -(y_{\text{min},t} + y_{\text{translation}}))$, where $y_{\text{min},t}$ is the minimum y-coordinate of all joint centers at time step t, and $y_{\text{translation}}$ is the estimated ground height.
- **Floating**: Displacement above the ground at each time step for all subjects which is defined as: $\text{Floating}(\overline{F}) = \frac{1}{T}\sum_{t=1}^{T} \max(0, y_{\text{min},t} + y_{\text{translation}})$
- **Sliding**: Measures the velocity of the lowest point in contact with the ground at consecutive time steps, $\text{Sliding}(\overline{S}) = \frac{1}{T-1}\sum_{t=1}^{T-1} \|\mathbf{p}_{\text{low},t+1} - \mathbf{p}_{\text{low},t}\|$, where $\mathbf{p}_{\text{low},t}$ is the 3D position of the lowest joint center at time step t.

4.4. Quantitative Results

We present a quantitative comparison of BIGE with other baseline models in Table 2 and Table 3. Table 2 evaluates the models using the metrics specified in Section 4.3. BIGE achieves the lowest fidelity error among all models, including MoCap data, and demonstrates diversity closest to the reference data, indicating that its reconstructions closely approximate clinician-approved real motions.

For the penetration and floating metrics, BIGE performs comparably to other baselines but shows significant improvement in the 'sliding' metric. This highlights that the motions generated by BIGE are more grounded, resulting in increased stability and naturalness compared to other models.

Table 3 compares the models based on the biomechanical constraints introduced in Section 3.3. BIGE achieves the best results across most categories, demonstrating that it generates the most biomechanically accurate motions. This indicates that BIGE effectively adheres to clinician-accepted constraints, highlighting its efficacy in producing physiologically meaningful motion.

4.5. Qualitative Comparison

We provide a frame-by-frame comparison of BIGE, MDM and VQVAE over a single squat cycle in Figure 3 and Figure 4 (limited to 5 frames for interpretability). The motions generated by MDM and VQVAE fail to achieve sufficient depth in the squat and exhibit unnatural artifacts during the cycle. In contrast, BIGE produces motions closely aligned with the reference data and free of such artifacts. The biomechanical constraints enforced in BIGE enable the model to perform deep, natural squats throughout the cycle. Comparisons with other baselines are provided in the supplementary material.

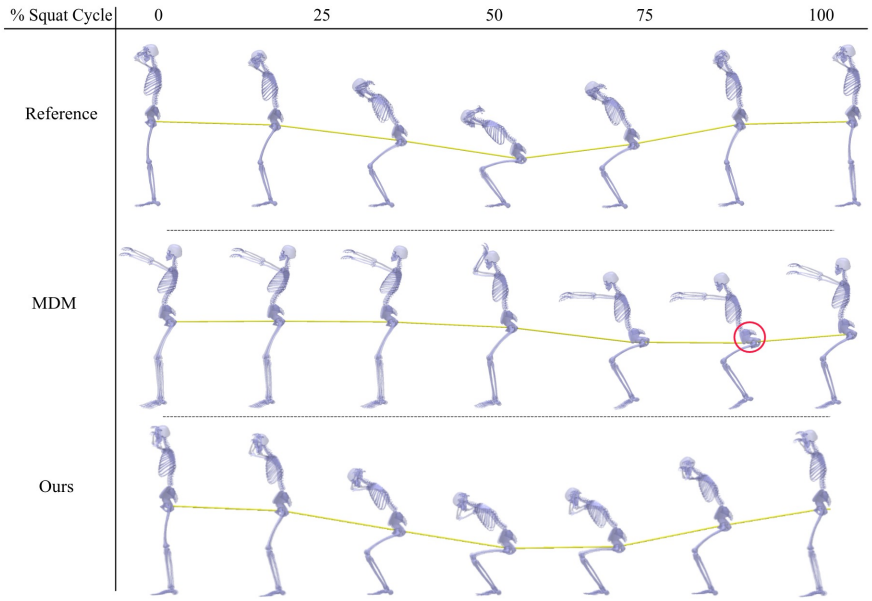

Figure 3: **Qualitative Results:** Comparison of generated samples from MDM and BIGE with the reference data. The yellow curve represents the movement of the hip joint over the entire squat cycle. BIGE generates a more natural squat motion compared to MDM. The red circle highlights the artifact observed in the pelvic tilt for MDM-generated motion.

5. Discussion

In this work, we proposed a novel framework, BIGE, for generating physiologically meaningful human motions tailored to clinician-defined objectives. By combining a VQ-VAE with a guidance mechanism based on biomechanical constraints, our approach generates motions that satisfy both observable kinematic properties and hidden criteria, such as muscle activations. This integration

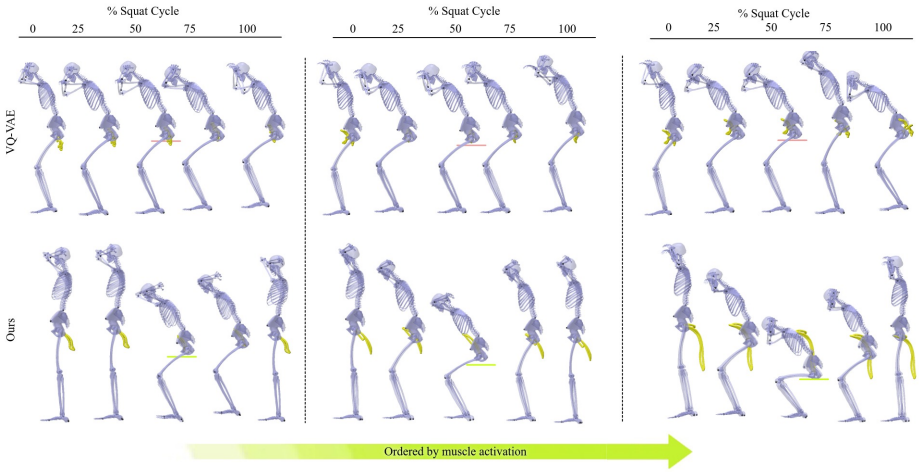

Figure 4: **Ablation:** Comparison between generated samples from VQVAE and our guidance strategy. The samples are ordered by the peak muscle activation. The red and green lines at 50% squat cycle represent the depth of the squat. Our guidance strategy leads to a more physiologically accurate squat motion as evidenced by the increased depth of the squat as muscle activation increases.

ensures that the generated motions are not only anatomically plausible but also clinically relevant for applications such as rehabilitation and exercise training.

Our results highlight the efficacy of BIGE in generating high-quality motions compared to state-of-the-art models. Quantitative evaluations demonstrate that BIGE achieves the closest fidelity and diversity to reference data while maintaining superior adherence to biomechanical constraints. Specifically, BIGE significantly reduces sliding artifacts and achieves smoother, more grounded motions than baseline models. Qualitative comparisons further show BIGE's ability to produce deep and natural squats free from artifacts, outperforming other generative models like MDM.

Despite its advantages, BIGE has some limitations that open avenues for future research. First, the current framework relies on pre-defined scoring metrics and constraints, which may not generalize well to diverse motions or varying clinical requirements. Future work could explore adaptive scoring mechanisms that learn from clinician feedback or dynamically adjust constraints based on specific use cases. Second, BIGE primarily focuses on generating motions within the joint kinematics framework. Incorporating additional biomechanical factors, such as EMG data and external forces, could further enhance the physiological validity of the generated motions. Exploring these aspects could also expand the framework's applicability to more complex motions and tasks.

6. Conclusion

Searching for optimal body motions based on individual variations in physical conditions can significantly assist athletes and clinicians. Our work recommends specific joint angle modifications that athletes can implement to improve their form. BIGE allows experts to accommodate these individual features to provide the most optimal way to execute a motion, thereby notably improving athletic performance. This work has significant implications for long-term outcomes such as ACL rehabilitation and injury prevention. In this paper, we identified and addressed the bottlenecks associated with using musculoskeletal models for human motion generation. Through a comprehensive exploration of skeletal models, muscle dynamics, and optimization frameworks, our methodology promises to unlock new frontiers in understanding and generating human motion.

References

Friedl De Groote, Allison Kinney, Anil Rao, and Benjamin Fregly. Evaluation of direct collocation optimal control problem formulations for solving the muscle redundancy problem. *Annals of Biomedical Engineering*, 44, 03 2016. doi: 10.1007/s10439-016-1591-9.

Scott L. Delp, Frank C. Anderson, Allison S. Arnold, Peter Loan, Ayman Habib, Chand T. John, Eran Guendelman, and Darryl G. Thelen. Opensim: Open-source software to create and analyze dynamic simulations of movement. *IEEE Transactions on Biomedical Engineering*, 54(11): 1940–1950, 2007. doi: 10.1109/TBME.2007.901024.

Christopher L. Dembia, Nicholas A. Bianco, Antoine Falisse, Jennifer L. Hicks, and Scott L. Delp. Opensim moco: Musculoskeletal optimal control. *PLOS Computational Biology*, 16(12):1–21, 12 2021. doi: 10.1371/journal.pcbi.1008493. URL https://doi.org/10.1371/journal.pcbi.1008493.

Peter Eckmann, Kunyang Sun, Bo Zhao, Mudong Feng, Michael K Gilson, and Rose Yu. Limo: Latent inceptionism for targeted molecule generation. 2022.

Antoine Falisse, Gil Serrancolí, Christopher L Dembia, Joris Gillis, and Friedl De Groote. Algorithmic differentiation improves the computational efficiency of opensim-based trajectory optimization of human movement. *PLoS One*, 14(10):e0217730, 2019.

Chuan Guo, Xinxin Zuo, et al. Action2motion: Conditioned generation of 3d human motions. *ACMMM*, 2020.

Chuan Guo, Shihao Zou, Xinxin Zuo, Sen Wang, Wei Ji, Xingyu Li, and Li Cheng. Generating diverse and natural 3d human motions from text. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5152–5161, 2022.

Debtanu Gupta, Shubh Maheshwari, Sai Shashank Kalakonda, Manasvi, and Ravi Kiran Sarvadevabhatla. Dsag: A scalable deep framework for action-conditioned multi-actor full body motion synthesis. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, January 2023.

Archibald Vivian Hill. The heat of shortening and the dynamic constants of muscle. *Proceedings of the Royal Society of London. Series B - Biological Sciences*, 126(843):136–195, 1938. doi: 10.1098/rspb.1938.0050. URL https://royalsocietypublishing.org/doi/abs/10.1098/rspb.1938.0050.

Sai Shashank Kalakonda, Shubh Maheshwari, and Ravi Kiran Sarvadevabhatla. Action-gpt: Leveraging large-scale language models for improved and generalized action generation. In *arXiv preprint https://arxiv.org/abs/2211.15603*, 2022.

Marilyn Keller, Keenon Werling, Soyong Shin, Scott Delp, Sergi Pujades, C. Karen Liu, and Michael J. Black. From skin to skeleton: Towards biomechanically accurate 3d digital humans. In *ACM ToG, Proc. SIGGRAPH Asia*, volume 42, December 2023.

Adrian Lai, Allison Arnold, and James Wakeling. Why are antagonist muscles co-activated in my simulation? a musculoskeletal model for analysing human locomotor tasks. *Annals of Biomedical Engineering*, 45, 09 2017. doi: 10.1007/s10439-017-1920-7.

Sung-Hee Lee, Eftychios Sifakis, and Demetri Terzopoulos. Comprehensive biomechanical modeling and simulation of the upper body. *ACM Trans. Graph.*, 28(4), September 2009. ISSN 0730-0301. doi: 10.1145/1559755.1559756. URL https://doi.org/10.1145/1559755.1559756.

Matthew Loper, Naureen Mahmood, et al. Smpl: A skinned multi-person linear model. *SIGGRAPH Asia*, 34(6):248:1–248:16, 2015.

Shubh Maheshwari, Debtanu Gupta, and Ravi Kiran Sarvadevabhatla. Mugl: Large scale multi person conditional action generation with locomotion. In *WACV*, 2022.

Alexander Mordvintsev, Christopher Olah, and Mike Tyka. Inceptionism: Going deeper into neural networks. *Google research blog*, 20(14):5, 2015.

Mathis Petrovich, Michael J. Black, and Gül Varol. Action-conditioned 3D human motion synthesis with transformer VAE. In *ICCV*, 2021.

Apoorva Rajagopal, Christopher L. Dembia, Matthew S. DeMers, Denny D. Delp, Jennifer L. Hicks, and Scott L. Delp. Full-body musculoskeletal model for muscle-driven simulation of human gait. *IEEE Transactions on Biomedical Engineering*, 63:2068–2079, 2016. URL https://api.semanticscholar.org/CorpusID:3798525.

Ajay Seth, Jennifer L. Hicks, Thomas K. Uchida, Ayman Habib, Christopher L. Dembia, James J. Dunne, Carmichael F. Ong, Matthew S. DeMers, Apoorva Rajagopal, Matthew Millard, Samuel R. Hamner, Edith M. Arnold, Jennifer R. Yong, Shrinidhi K. Lakshmikanth, Michael A. Sherman, Joy P. Ku, and Scott L. Delp. Opensim: Simulating musculoskeletal dynamics and neuromuscular control to study human and animal movement. *PLOS Computational Biology*, 14(7):1–20, 07 2018. doi: 10.1371/journal.pcbi.1006223. URL https://doi.org/10.1371/journal.pcbi.1006223.

Guy Tevet, Sigal Raab, Brian Gordon, Yonatan Shafir, Daniel Cohen-Or, and Amit H Bermano. Human motion diffusion model. *arXiv preprint arXiv:2209.14916*, 2022.

Scott Uhlrich, Antoine Falisse, Łukasz Kidziński, Julie Muccini, Michael Ko, Akshay Chaudhari, Jennifer Hicks, and Scott Delp. Opencap: 3d human movement dynamics from smartphone videos. 07 2022. doi: 10.1101/2022.07.07.499061.

Aaron van den Oord, Oriol Vinyals, and Koray Kavukcuoglu. Neural discrete representation learning. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 6309–6318, Red Hook, NY, USA, 2017. Curran Associates Inc. ISBN 9781510860964.

Keenon Werling, Dalton Omens, Jeongseok Lee, Ioannis Exarchos, and Karen Liu. Fast and feature-complete differentiable physics engine for articulated rigid bodies with contact constraints. 07 2021. doi: 10.15607/RSS.2021.XVII.034.

Heyuan Yao, Zhenhua Song, Baoquan Chen, and Libin Liu. Controlvae: Model-based learning of generative controllers for physics-based characters. *ACM Transactions on Graphics*, 41(6):1–16, November 2022. ISSN 1557-7368. doi: 10.1145/3550454.3555434. URL http://dx.doi.org/10.1145/3550454.3555434.

Ye Yuan, Jiaming Song, Umar Iqbal, Arash Vahdat, and Jan Kautz. Physdiff: Physics-guided human motion diffusion model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 16010–16021, 2023.

Felix E. Zajac. Muscle and tendon: properties, models, scaling, and application to biomechanics and motor control. *Critical reviews in biomedical engineering*, 17 4:359–411, 1989. URL https://api.semanticscholar.org/CorpusID:25888803.

Jianrong Zhang, Yangsong Zhang, Xiaodong Cun, Shaoli Huang, Yong Zhang, Hongwei Zhao, Hongtao Lu, and Xi Shen. T2m-gpt: Generating human motion from textual descriptions with discrete representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023a.

Mingyuan Zhang, Huirong Li, Zhongang Cai, Jiawei Ren, Lei Yang, and Ziwei Liu. Finemogen: Fine-grained spatio-temporal motion generation and editing. *NeurIPS*, 2023b.