

Problem

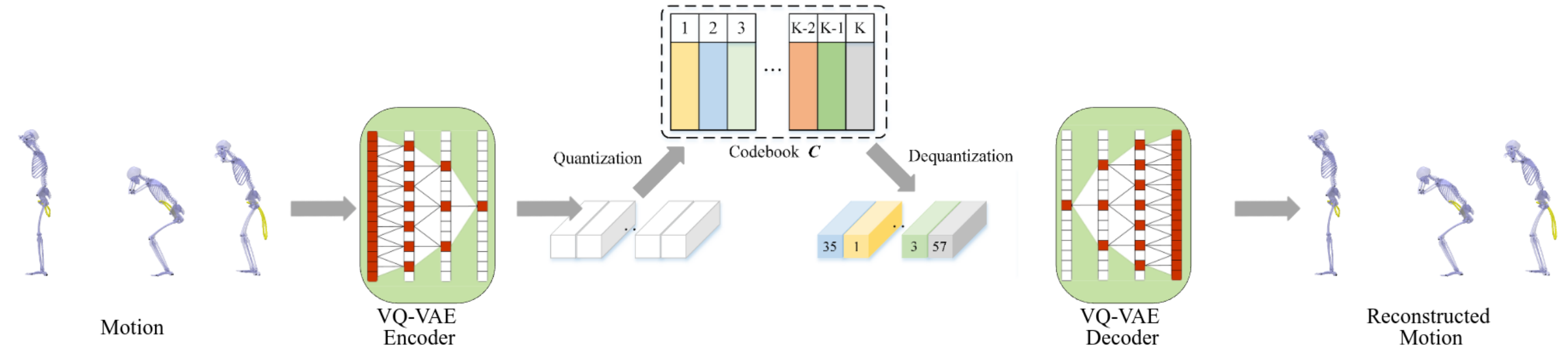
How to guide generative models to generate biomechanically accurate human motion?

Our Results

We propose a novel framework, *BIGE*, that combines bio-mechanically meaningful scoring metrics with generative modeling. To achieve this, we:

- develop and validate crucial physiological scoring metrics for injury avoidance and rehabilitation. Our scores incorporate observable and hidden biomechanical constraints, ensuring the clinical relevance of generated motions.
- integrate a differentiable surrogate model for muscle activation to reverse optimize the latent space of the generative model:
- enable the retrieval of physiologically valid motions through targeted search;
- demonstrates superior performance in generating diverse, physically plausible motions while maintaining high fidelity to clinician-defined objectives compared to existing approaches.
- We open source the code for simulation, data, and models at:
Code: <https://github.com/Rose-STL-Lab/BIGE>

Generative Model



We train a generative model based on Vector Quantized Variational Autoencoder (VQVAE[3]) for learning diverse motion representations of the joint kinematics.

Biomechanical Constraints

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, joint center dynamics, and muscle activations.

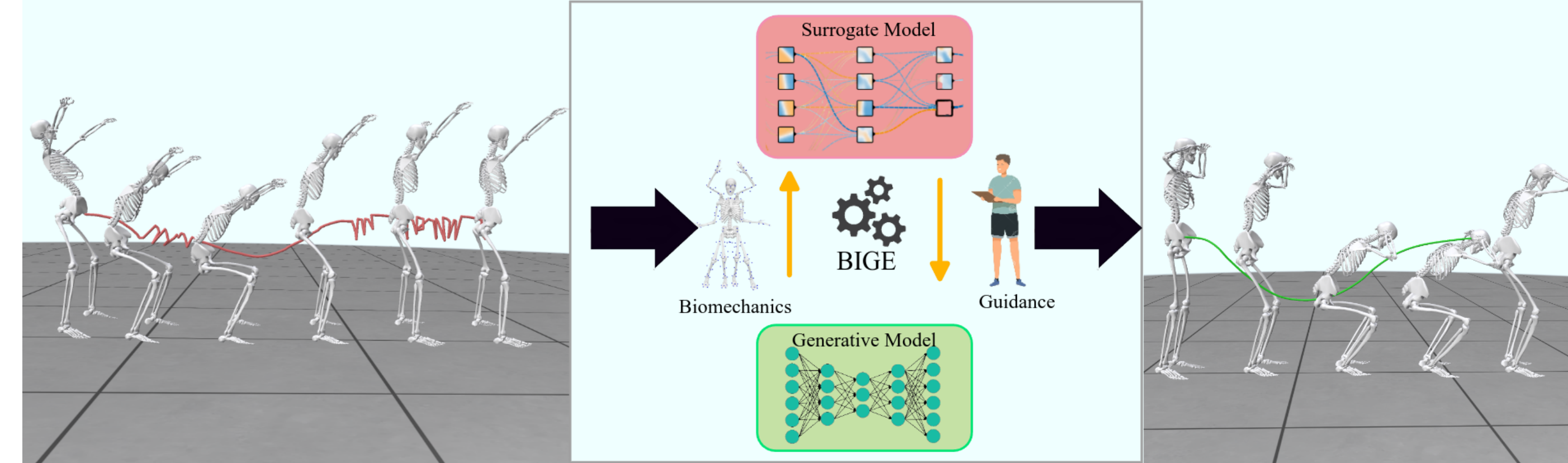
Metric	Definition	Guidance Loss
Pelvis Tilt (degs)	$\max(K_{\text{pelvis}})$, where K_{pelvis} is the pelvis tilt angle at peak activation	$\mathcal{L}_{\text{tilt}} = \frac{1}{T} \sum_{t=1}^T \hat{K}_{t, \text{peak_timestep.pelvis_tilt_index}}$
Asymmetry (degs)	$\text{mean}(\sum_{(j_l, j_r) \in \text{leg joints}} K_{j_l, \text{left}} - K_{j_r, \text{right}})$, comparing the angles of left and right leg joints	$\mathcal{L}_{\text{asym}} = \frac{1}{T} \sum_{t=1}^T \sum_{(j_l, j_r) \in \text{leg joints}} (\hat{K}_{t, \text{left}} - \hat{K}_{t, \text{right}})^2$
Angular vel. (degs/s)	$\text{mean}(\sum_{\theta \in K^{\text{angle}}} \frac{d\theta}{dt})$, where K^{angle} is joint angles in Kinematics	$\mathcal{L}_{\omega} = \frac{1}{T} \sum_{t=1}^T \sum_{\theta \in K^{\text{angle}}} (\hat{K}_{t, \theta} - \hat{K}_{t, \theta})^2$
COM vel. (m/s)	$\text{mean}(\ \frac{d\hat{K}^{\text{trans}}}{dt}\)$, where $K^{\text{trans}} \in \mathbb{R}^3$ is the pelvis translation relative to the ground	$\mathcal{L}_{v\text{COM}} = \frac{1}{T-1} \sum_{t=1}^{T-1} (\hat{K}_{t+1}^{\text{trans}} - \hat{K}_t^{\text{trans}})^2$
COM acc. (m/s ²)	$\text{mean}(\ \frac{d^2\hat{K}^{\text{trans}}}{dt^2}\)$, where ∇^2 is the Laplacian operator	$\mathcal{L}_{a\text{COM}} = \frac{1}{T-2} \sum_{t=2}^{T-2} (\nabla^2 \hat{K}^{\text{trans}})^2$
Foot Sliding	$\text{mean}(\ \frac{d\mathbf{p}_{\text{lowest}}}{dt}\)$, ensures the lowest point $\mathbf{p}_{\text{lowest}}$ 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: Summary of loss constraints. The final objective is formulated as a weighted sum of the individual constraints.

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 [2]	−28.13±6.03 2.76±0.59 96.24±13.96 0.67±0.14 30.45±5.55 0.03±0.03
MDM [1]	0.42±1.87 2.25±0.28 554.10±112.77 3.39±0.79 335.86±81.37 0.03±0.02
T2M-GPT [4]	4.16±2.42 4.73±0.29 918.14±42.06 6.09±0.48 597.95±34.60 0.02±0.01
VQVAE [3]	−20.60±0.51 1.12±0.03 37.78±0.42 1.59±0.08 128.24±6.49 0.16±0.01
BIGE	−23.09±1.54 2.09±0.14 56.95±1.94 0.51±0.04 27.32±3.88 0.06±0.01

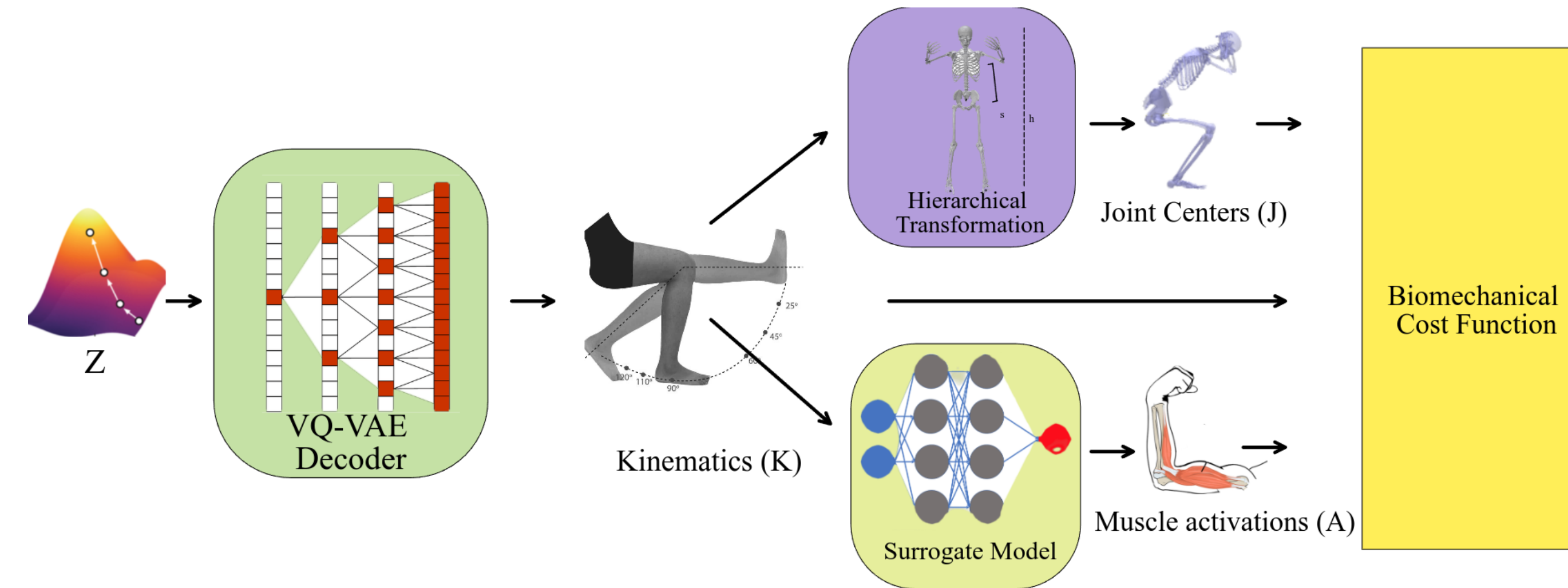
Table: Performance evaluation on guidance metrics. The right arrow (→) indicates that the values should be close to the Reference data. Down arrow (↓) means a value closer to 0 is better. BIGE effectively adheres to clinician-accepted constraints.

Score-Guided Generative Model



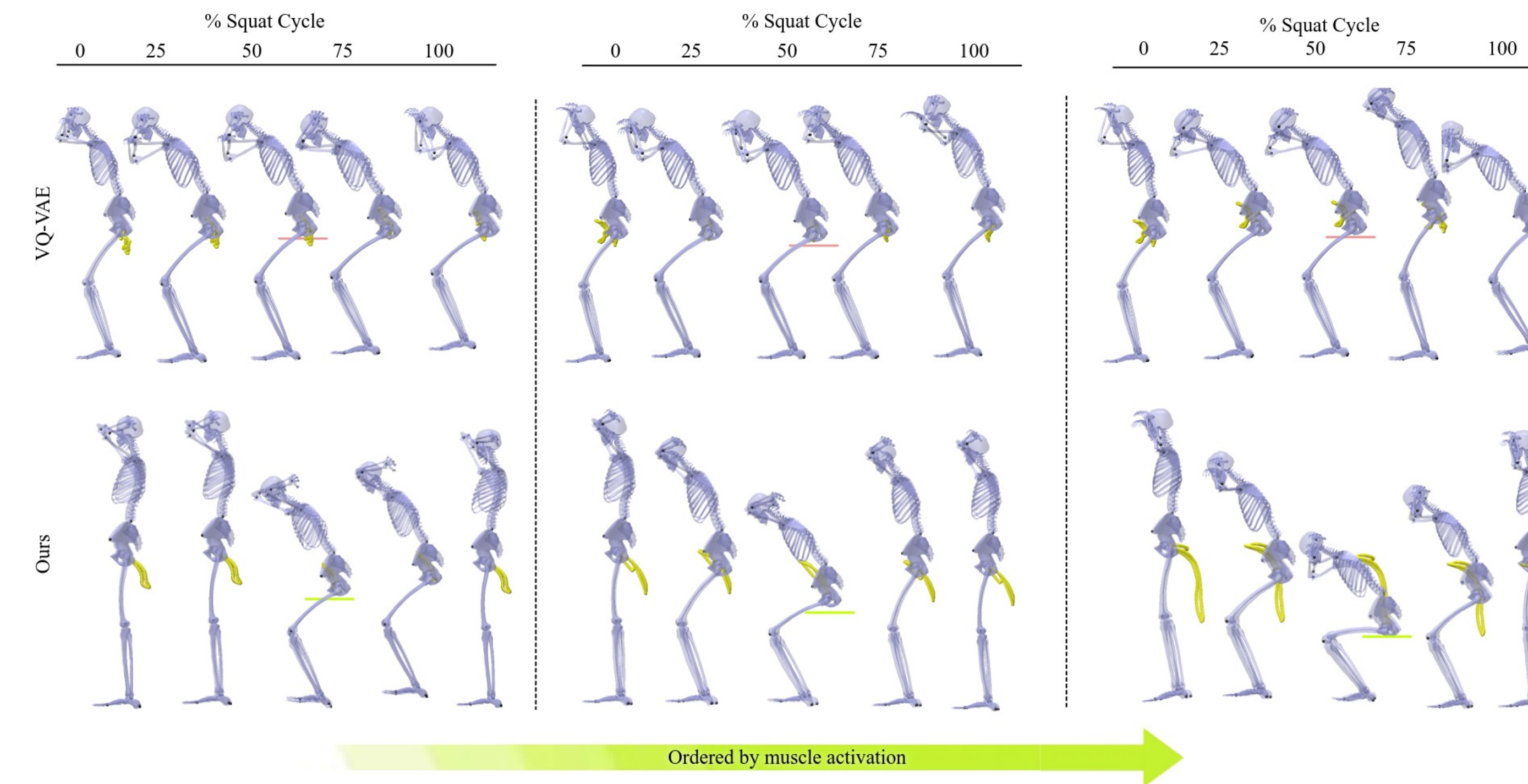
BIGE leverages score-guided optimization to refine latent representations in generative models, enforcing biomechanical constraints to produce high-quality human motion sequences.

Score calculation



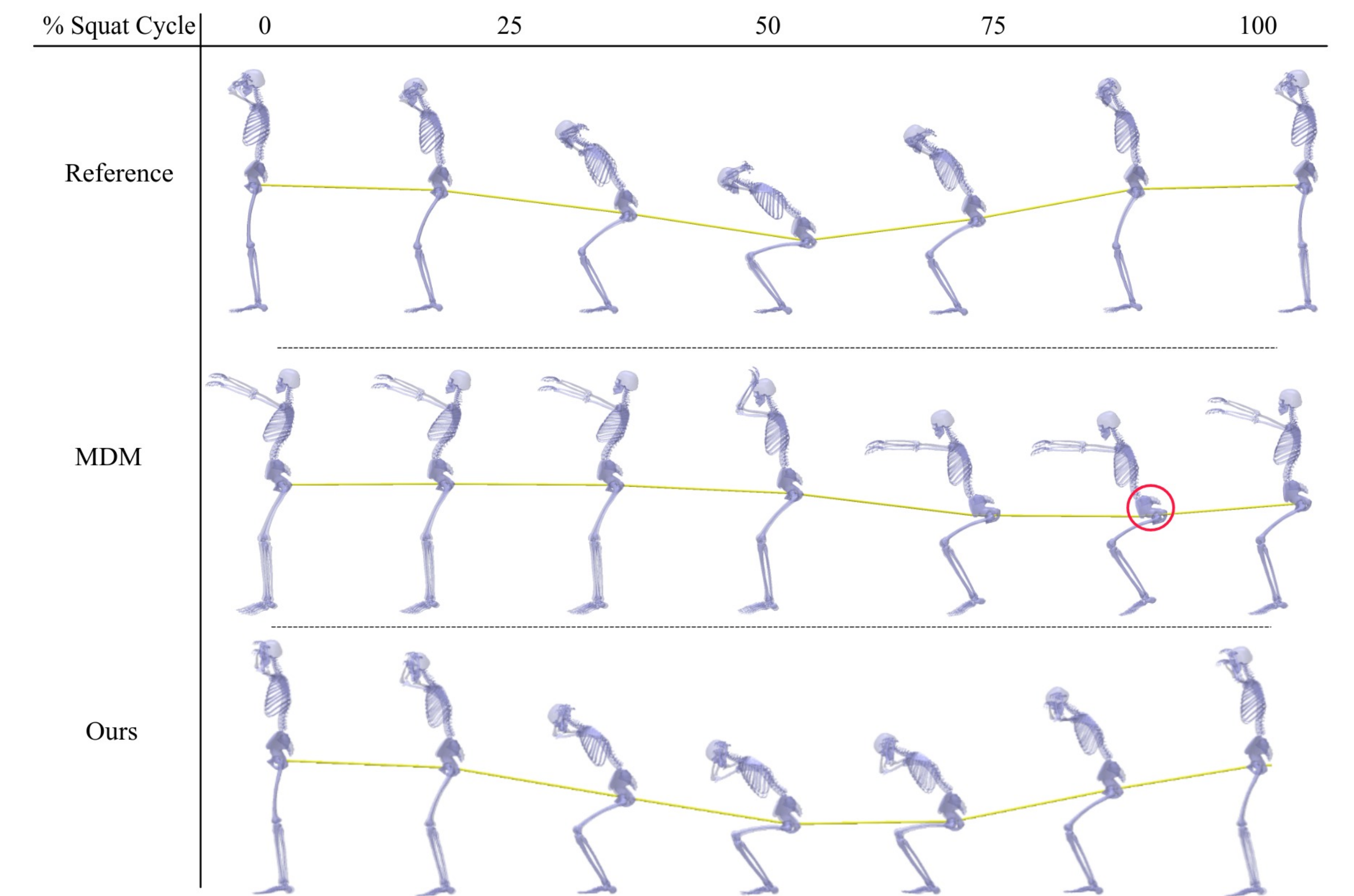
Latent variables (sampled randomly) are decoded using the VQ-VAE decoder to generate joint kinematics. Next, hierarchical transformations are applied to the biomechanical model, to compute joint centers and a surrogate model to predict muscle activations. Finally, clinician-defined constraints are imposed on the derived variables.

Muscle activation guidance



Our guidance strategy leads to a more physiologically accurate squat motion as evidenced by the increased depth of the squat as muscle activation increases. The samples are ordered by the peak muscle activation. The red and green lines at 50% squat cycle represent the depth of the squat.

Qualitative Comparison



Comparison of squat motion sequences generated by MDM [1] and BIGE against reference data from OpenCap [2]. The yellow trajectory traces hip joint movement throughout the complete squat cycle. BIGE produces biomechanically accurate squats that closely match the reference, while MDM exhibits unrealistic pelvic tilt (highlighted in red), resulting in less natural movement.

Quantitative Comparison

Model	Realism		Floor metrics		
	Fidelity ↓	Diversity →	Penetration (cm) ↓	Floating (cm) ↓	Sliding ($\frac{m}{sec}$) ↓
Reference	-	1.736	0.003±0.057	0.000±0.021	0.035±0.024
MoCap [2]	7.855±9.902	1.773±0.015	0.501 ±0.987	0.759 ±1.649	0.209 ±0.161
MDM [1]	257.95±105.92	1.26±0.05	2.685±3.840	3.216±4.636	1.395±2.672
T2M-GPT [4]	459.49±257.30	1.45±0.05	1.997±2.938	2.440±3.606	3.770±4.758
VQVAE [3]	45.99±4.67	1.82±0.03	2.035±3.016	2.331±3.506	1.351±0.830
BIGE	6.48 ±4.46	1.76 ±0.03	2.327±3.659	2.379±3.561	0.291±0.255

Table: 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.

References

- [1] G. Tevet, S. Raab, B. Gordon, Y. Shafir, D. Cohen-Or, and A. H. Bernano. Human motion diffusion model. *arXiv preprint arXiv:2209.14916*, 2022.
- [2] S. Uhlich, A. Falisse, Kidziński, J. Muccini, M. Ko, A. Chaudhari, J. Hicks, and S. Delp. Opencap: 3d human movement dynamics from smartphone videos. 07 2022.
- [3] A. van den Oord, O. Vinyals, and K. 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.
- [4] J. Zhang, Y. Zhang, X. Cun, S. Huang, Y. Zhang, H. Zhao, H. Lu, and X. 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)*, 2023.

Acknowledgments

We extend our gratitude to The Wu Tsai Human Performance Alliance for their invaluable support and resources. Their dedication to advancing human performance through interdisciplinary research has been instrumental to our project.