

Modeling Kinematic Signatures of Aging for Human Motion Prediction in Assistive Robotics

Yiran Liu^{1,2}, Eugene Zhao^{1,2}, Karim Habashy^{1,2}, Joanna Peng^{1,2}, Jason Peng³, Chang Shu¹, Pengcheng Xi¹

¹National Research Council Canada ²University of Waterloo ³Simon Fraser University

y3275liu@uwaterloo.ca, chang.shu@nrc-cnrc.gc.ca, pengcheng.xi@nrc-cnrc.gc.ca

Abstract—We present an age-conditioned human motion generation framework for studying aging signatures in the context of assistive robotics. Our pipeline converts clinical gait recordings into learning-ready motion representations while addressing defects from legacy motion-capture data. Using 420 clips from 138 subjects across young, mid-age, and elderly groups, we train a Spatial-Temporal Graph Convolutional Network (ST-GCN) to recognize age group from motion clips, achieving 64.6% classification accuracy. We then fine-tune a Motion Diffusion Model (MDM) with Low-Rank Adaptation (LoRA) to generate text-conditioned human motion with age-group control, and compare against the unadapted MDM under both neutral and elderly-described text prompts. To evaluate whether generated motions preserve aging-related characteristics, we analyze clinical and kinematic gait metrics on 9,000 generated clips. Our results show that the LoRA-adapted model reproduces spatial kinematic constraints in the Elderly condition — a signature the unadapted MDM fails to express even under elderly-described prompts — but spatiotemporal consistency remains unresolved, with generated stride variability massively overshooting clinical baselines, which we attribute to the capacity ceiling of low-rank, discrete-token adaptation. These findings highlight dataset size, motion quality, and conditioning strength as key bottlenecks, and motivate larger age-diverse datasets for demographic-aware motion prediction and assistive robot planning. Code and video available on the [project page](#).

I. INTRODUCTION

Accurate human motion anticipation is foundational for safe human-robot interaction, especially when robots share space with humans, plan around future movement, or provide proactive assistance. Recent deep generative architectures, particularly Motion Diffusion Models (MDMs) [1], [2], have demonstrated strong capability in synthesizing realistic and diverse human motions. However, widely used motion-capture datasets such as AMASS [3] are heavily biased toward young, healthy actors [4], so existing generative models cover diverse actions and styles [5] but rarely capture population-level biomechanical factors such as age [6]. This limitation is particularly critical for assistive robots in aging-in-place and eldercare scenarios, where anticipating older adults’ movements is essential for safe navigation, timely assistance, and proactive interaction.

Although MDMs are trained for synthesis rather than direct forecasting, they implicitly learn the distribution

over plausible future motions — precisely what a prediction system samples from under partial observation. Age-conditioning this distribution restricts samples to a demographic-specific manifold, giving downstream forecasters and robot planners priors aligned to the user population they serve. An age-aware generative prior is therefore a precursor to, rather than a substitute for, motion prediction in eldercare contexts; clinical gait analysis, which has long quantified spatiotemporal aging signatures [7], [8], provides the supervision signal needed to build it.

To address this gap, we present a proof-of-concept framework that integrates clinical gait analysis with deep generative modeling toward age-aware human motion synthesis. We develop a data processing pipeline that converts clinical gait recordings into standardized skeletal representations, fine-tune a Spatial-Temporal Graph Convolutional Network (ST-GCN) [9] to infer age group and encode age-related biomechanical characteristics into a latent representation, and train a Low-Rank Adaptation (LoRA) [10] adapter for an MDM to generate text-driven motions conditioned on age group. We evaluate generated motions with clinical and kinematic gait metrics and against the unadapted MDM under both neutral and elderly-described prompts. Our analysis shows that the adapted model captures spatial kinematic constraints (reduced hip Range-of-Motion in the Elderly condition) that the unadapted MDM fails to express even when prompted explicitly for old age, but fails to reproduce spatiotemporal consistency: stride variability overshoots clinical baselines, a limitation we attribute to the capacity ceiling of low-rank, discrete-token adaptation. These findings motivate larger, age-diverse human motion datasets and stronger conditioning mechanisms for downstream motion prediction and assistive robot planning.

II. METHODOLOGY: AGE-AWARE MOTION SYNTHESIS FRAMEWORK

A. Data Processing Pipeline

We adopt the motion capture dataset of Van Crielinge et al. [11] (ages 21–86) and convert clinical marker recordings into HumanML3D-format skeletal graphs [12] via SMPL body-model fitting [3] and feature extraction, as summarized in Fig. 1 (top). As shown in Fig. 2, naive mapping causes mesh inflation (right), while standard priors induce crouch artifacts (left); our multi-stage optimization pipeline adapted from VPoser [13] addresses both to recover a natural posture

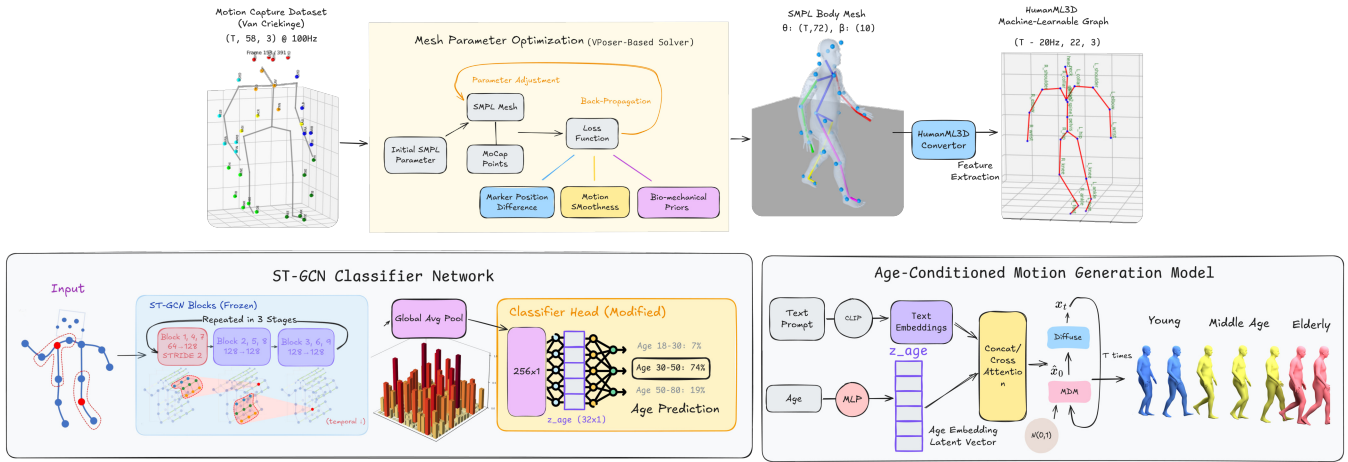


Fig. 1: Overview of our age-aware motion synthesis framework. *Top*: clinical marker recordings are converted into the 263-dimensional HumanML3D representation via VPoser-regularized SMPL fitting followed by feature extraction. *Bottom*: an ST-GCN++ age representation learner extracts \mathbf{z}_{age} (left) and an age-conditioned Motion Diffusion Model (right) generates text-driven motion.

(center) and resolves common clinical-dataset defects such as axis misalignment, marker mislabeling, and foot contact issues.

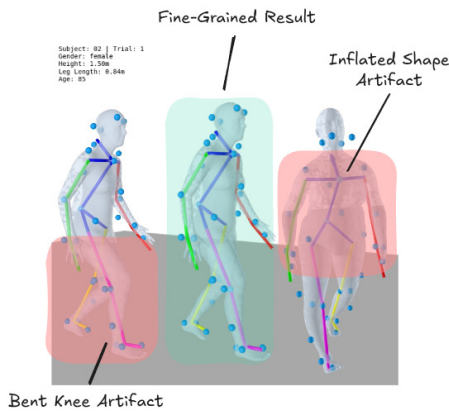


Fig. 2: Artifact analysis in clinical motion retargeting.

B. Age Signature Detecting Network

As illustrated in Fig. 1 (bottom left), an Age Representation Learner built on an ST-GCN [9] is fine-tuned to classify motion clips into three age groups (Young, Mid, Elderly) and regress biological age. Once the network can successfully predict age, the penultimate layer encapsulates the biomechanical signatures of aging (e.g., joint motion range). We extract this vector to serve as the age latent embedding, denoted as \mathbf{z}_{age} .

C. Age-Conditioned Generative Network

Our target framework, shown in Fig. 1 (bottom right), is a Conditional Diffusion Model [1], [2] that denoises a latent motion representation \mathbf{x}_T conditioned on both text and age. At training time, \mathbf{z}_{age} extracted by the frozen ST-GCN encoder is injected via cross-attention, while a small

MLP $g(a)$ is jointly trained to map a scalar age to the same latent space, enabling continuous age-controlled generation at inference from a desired age value alone.

Training this full architecture from scratch is infeasible under our compute and dataset budget, so the experiments in this paper validate a discrete-token prototype instead: we apply Low-Rank Adaptation (LoRA) [10] to a pre-trained MDM with three style tokens (young, mid, old) and treat the continuous-conditioning architecture as the natural next step rather than a delivered contribution. This honest scoping isolates the bottlenecks that future continuous conditioning must overcome.

D. Gait-Analysis Metrics for Evaluation

In order to validate that the aging signal is preserved in the processed dataset and successfully reproduced in the generated motion clips, we adapted several standard, training-free gait-analysis metrics documented in clinical literature [14], [15], [8].

- **Spatiotemporal Parameters:** We compute average walking speed, stride length, and cadence by detecting heel-strike events based on the zero-velocity crossings of the ankle joints.
- **Gait Variability:** We calculate the Coefficient of Variation (CV) for stride length and stride time across multiple steps.
- **Joint Range of Motion (RoM):** We measure the peak flexion and extension angles of the hip and knee joints during the gait cycle.

These metrics are computed deterministically from sample clips within our dataset and from motion generated by our LoRA-based prototype, providing a quantitative, biomechanically grounded evaluation of our framework.

III. EXPERIMENTS AND RESULTS

A. Data Quality Validation

The Van Criekinge et al. [11] clinical motion capture dataset contains recordings from 138 subjects across the lifespan. During preprocessing, we resolved dataset-specific artifacts including sensor measurement errors, initialization transients, and mislabeled markers; after discarding 138 T-pose calibration clips, 24 short clips (< 40 frames), and 4 manually identified artifact clips, the data loader yields 420 effective walking clips (Young 115 / Mid-age 154 / Elderly 151). To preserve biological fidelity during conversion to the HumanML3D machine-learnable format, we applied a multi-stage optimization procedure adapted from VPoser [13]: the optimizer first recovers global body pose, then refines localized artifacts around body shape and limbs. The fitting loss converges rapidly and eliminates the mesh-inflation and crouch-walking artifacts of Fig. 2; the converged residual is non-zero, reflecting legacy marker layout and remaining occlusion errors that contribute to the within-group variance visible in our distributional comparisons.

B. Model Training Results

1) *ST-GCN Age Representation Learner*: We fine-tuned the ST-GCN++ architecture to classify the processed walking clips into three age categories: Young (< 35), Mid-age (35–60), and Elderly (≥ 60).

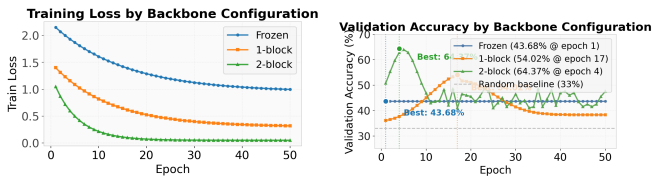


Fig. 3: Training dynamics for the ST-GCN age classifier. Left: training loss. Right: validation accuracy.

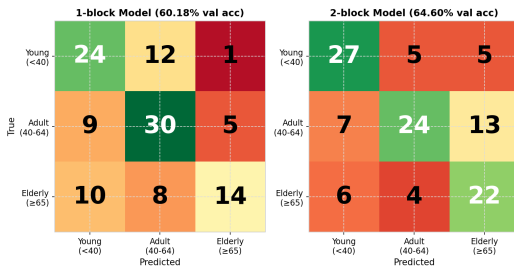


Fig. 4: Confusion matrices for the ST-GCN predictions.

These boundaries are not arbitrary: they roughly balance clip counts across bins and align with clinical findings that the kinematic gap between Young and Mid-age is small relative to within-group variance, whereas distinguishable spatiotemporal decline becomes consistent past 60 [16]. To prevent subject-based overfitting — where the model memorizes individual gait patterns — we ensured strict, mutually exclusive subject assignment between training and

validation, with an approximate 3:1 clip split (315 train / 105 validation clips, subject-disjoint across the 138 subjects). We progressively unfreeze backbone blocks (Fig. 3) to balance validation accuracy (64.60%) against overfitting, which remains visible as a gap between validation accuracy and the near-perfect training fit. The confusion matrices (Fig. 4) show that the model consistently separates the kinematic extremes (Young vs. Elderly), with the primary failure mode being the fragmentation of the Mid-age category — a feature, not only a bug, given the small Young \leftrightarrow Mid kinematic gap reported in clinical studies. We therefore report quantitative trends as Young \rightarrow Elderly contrasts and include Mid-age as a smooth interior point rather than a hard binary boundary.

2) *Generative Motion Synthesis*: As a proof of concept, we conditioned the LoRA-adapted MDM using discrete age tokens. After training LoRA adapters at ranks 5, 12, and 16, we selected the lowest-loss adapter (rank 12) and generated 3,000 clips per age condition (9,000 total) from the styled prompt template “A person is walking forward in {young/mid/old} style”, with classifier-free guidance scale $s = 1.0$ to remain within the conditional distribution learned during LoRA training. To isolate the effect of the adapter, we additionally generated baseline clips from the *unadapted* MDM using a neutral prompt “A person is walking forward.” and an explicit-age prompt “An old person is walking forward.”.

C. Quantitative Clinical Analysis

We computed standard clinical metrics (Spatiotemporal Parameters, Gait Variability, and Joint Range of Motion) on both the processed ground-truth dataset and the 9,000 generated motion clips.

1) *Dataset Baseline Validation*: Analysis of the ground-truth dataset confirms that the expected aging signal is preserved end-to-end through our conversion pipeline: monotonic Young \rightarrow Elderly trends consistent with clinical literature are visible for walking speed, stride length, knee RoM, and stride-time CV (Fig. 7), and per-clip linear regressions against chronological age (Fig. 6) reproduce these directions with weak-to-moderate correlations.

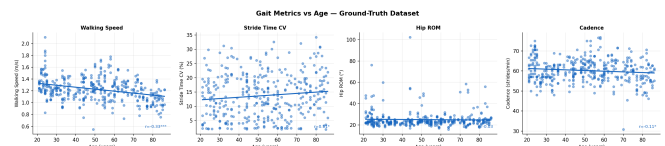


Fig. 6: Per-clip linear regression of clinical gait metrics against chronological age on the processed VC dataset.

2) *Generated Result*: Fig. 5 shows representative motion sequences generated from the LoRA-adapted MDM under each age token. Coarse stylistic cues of aging are visible: the Elderly sample exhibits a noticeably hunched back, and stride-to-stride foot-clearance height differs visibly between the Young and Elderly conditions. Visual inspection alone is, however, an unreliable proxy for the underlying biomechanics, motivating the quantitative analysis below.

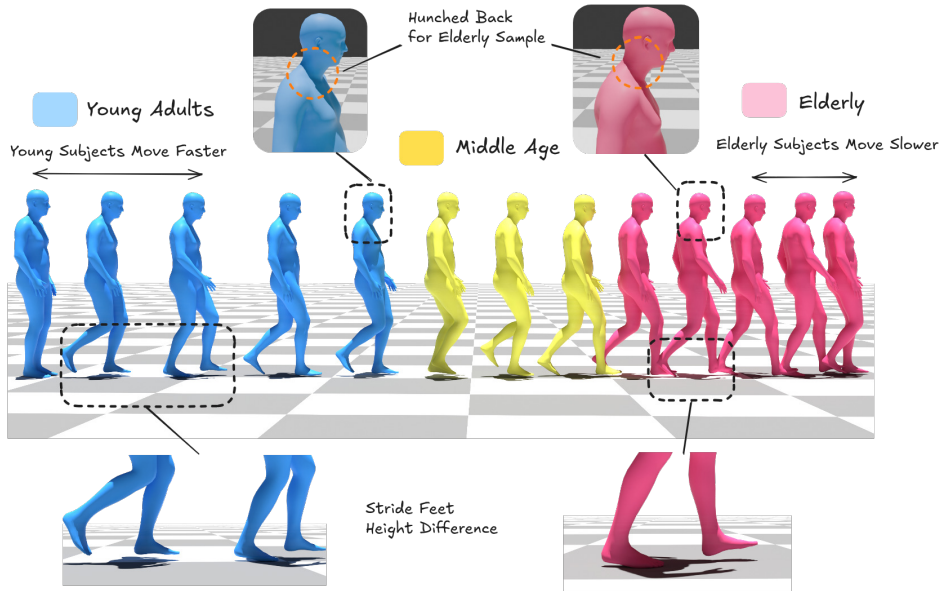


Fig. 5: Generated walking sequences across the three age tokens (young, mid, old). The Elderly sample shows a noticeably hunched back and reduced foot-clearance height relative to the Young sample.

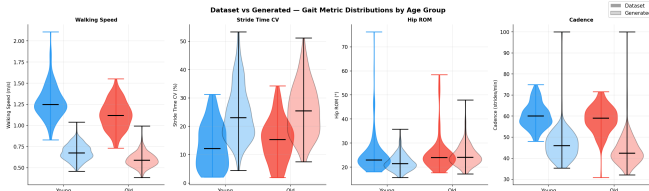


Fig. 7: Violin plots comparing the distribution of clinical metrics between the ground-truth dataset and the generated motion clips across age groups.

TABLE I: Welch two-sided t -test of Young–Old contrast on per-clip metrics at confidence level $C = 95\%$. Values in red are not statistically significant ($p \geq 0.05$).

Metric (Y–O)	Dataset Δ	Generated Δ
Walking Speed (m/s)	+0.155	+0.047
Stride-Time CV (%)	-1.87	-0.20
Hip RoM ($^\circ$)	-0.25	-0.65
Cadence (steps/min)	+1.61	+1.64
Joint Lin. Vel. (m/s)	+0.140	+0.046

3) *Quantitative Gait Metrics Comparison:* Table I reports Welch two-sided t -tests of the Young–Old contrast on per-clip metrics, on both the dataset and the 9,000 LoRA-generated clips. Generated motions reproduce the expected sign and significance for walking speed, cadence, and joint linear velocity; hip RoM becomes significant only in the generated condition ($p=0.019$), so the adapter sharpens spatial constraints rather than merely mirroring them — an effect the unadapted-MDM baselines (which fall within the young-token RoM range under both neutral and elderly-described prompts) cannot produce from text alone, confirming that the AMASS prior concentrates on young, healthy actors. Stride-time CV, however, remains non-significant and uniformly

elevated (16–19%) across all generated clips, far exceeding clinical reference values — a failure we attribute in the Discussion to the capacity ceiling of low-rank, discrete-token adaptation rather than to text conditioning.

IV. DISCUSSION AND CONCLUSION

Our findings reveal both the promise and the current limitations of data-driven models for age-related motion signatures. The ST-GCN successfully separates the kinematic extremes (Young vs. Elderly), confirming that motion clips contain discriminative biomechanical information related to age group. The LoRA-adapted MDM learns and imposes spatial kinematic constraints that the unadapted MDM cannot express even under elderly-described prompts, yet spatiotemporal consistency remains unresolved: generated stride variability massively overshoots clinical baselines. We attribute this remaining failure to the capacity ceiling of low-rank, discrete-token adaptation: a rank-12 LoRA update can push high-leverage spatial statistics such as hip RoM toward an age-specific mean, but cannot coherently re-shape stride timing, leaving the base MDM prior to dominate the temporal axis of the kinematic chain. This gap highlights a core challenge for assistive robotics in eldercare: visually plausible age-conditioned motion is not sufficient — prediction systems built on top of generative priors must preserve the biomechanical characteristics that affect anticipation, safety, and interaction.

Dataset scale and discrete-token conditioning capacity are the two principal bottlenecks; future work will train the continuous latent-conditioning architecture of Section II end-to-end on larger, age-diverse datasets, with scalar age regression and t-SNE visualization of \mathbf{z}_{age} as classifier-side next steps — moving past the LoRA capacity ceiling toward demographic-aware motion priors for assistive eldercare.

REFERENCES

- [1] G. Tevet, S. Raab, B. Gordon, Y. Shafir, and A. H. Bermano, "Human motion diffusion model," in *ICLR*, 2023.
- [2] X. Chen, B. Jiang, W. Liu, Z. Huang, B. Fu, T. Chen, J. Yu, and G. Yu, "Executing your commands via motion diffusion in latent space," *arXiv preprint arXiv:2212.04048*, 2022, arXiv:2212.04048.
- [3] N. Mahmood, N. Ghorbani, N. F. Troje, G. Pons-Moll, and M. J. Black, "AMASS: Archive of motion capture as surface shapes," in *ICCV*, 2019.
- [4] R. Umagami, L. Yue, X. Chu, R. Fukushima, T. Narita, Y. Mukuta, T. Takahata, J. Yang, and T. Harada, "Intend to move: A multimodal dataset for intention-aware human motion understanding," in *NeurIPS 2025 Datasets and Benchmarks Track*, 2025.
- [5] H. Sawdayee, C. Guo, G. Tevet, B. Zhou, J. Wang, and A. H. Bermano, "Dance like a chicken: Low-rank stylization for human motion diffusion," *arXiv preprint arXiv:2503.19557*, 2025.
- [6] T.-H. Liao, Y. Zhou, Y. Shen, C.-H. P. Huang, S. Mitra, J.-B. Huang, and U. Bhattacharya, "Shape my moves: Text-driven shape-aware synthesis of human motions," 2025. [Online]. Available: <https://arxiv.org/abs/2504.03639>
- [7] M. C. Díaz, R. Jiménez, C. Saavedra, J. P. Vásquez, R. Torres, M. A. Solís, D. Velandia, and R. Salas, "Evaluating the influence of physical activity on gait aging using multilevel machine learning," *IEEE Access*, 2025.
- [8] S. Galasso, "Development of a machine learning-based marker-less gait analysis system for clinical applications," Ph.D. dissertation, Università degli Studi di Cassino e del Lazio Meridionale, 2024.
- [9] S. Yan, Y. Xiong, and D. Lin, "Spatial temporal graph convolutional networks for skeleton-based action recognition," in *AAAI*, 2018.
- [10] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, "LoRA: Low-rank adaptation of large language models," in *ICLR*, 2022.
- [11] T. Van Crielinge, W. Saeys, S. Truijten, L. Vereeck, L. H. Sloot, and A. Hallemans, "A full-body motion capture gait dataset of 138 able-bodied adults across the life span and 50 stroke survivors," *Scientific Data*, vol. 10, no. 1, p. 852, 2023.
- [12] C. Guo, S. Zou, X. Zuo, S. Wang, W. Ji, X. Li, and L. Cheng, "Generating diverse and natural 3d human motions from text," in *CVPR*, 2022, pp. 5152–5161.
- [13] G. Pavlakos, V. Choutas, N. Ghorbani, T. Bolkart, A. A. A. Osman, D. Tzionas, and M. J. Black, "Expressive body capture: 3d hands, face, and body from a single image," 2019. [Online]. Available: <https://arxiv.org/abs/1904.05866>
- [14] C. A. Bailey, A. Mir-Orefice, T. K. Uchida, J. Nantel, and R. B. Graham, "Smartwatch-based prediction of single-stride and stride-to-stride gait outcomes using regression-based machine learning," *bioRxiv*, 2023.
- [15] M. C. Cordeiro, C. Cathain, M. Nascimento, and T. B. Rodrigues, "A constraint-based synthetic data generation methodology for gait stability assessment in older adults," *Frontiers in Sports and Active Living*, vol. 7, p. 1646146, 2025.
- [16] N. Herssens, E. Verbecque, A. Hallemans, L. Vereeck, V. Van Rompaey, and W. Saeys, "An investigation of the spatio-temporal parameters of gait and margins of stability throughout adulthood," *Journal of The Royal Society Interface*, vol. 17, no. 166, p. 20200194, 2020.