8th Long-term Human Motion Prediction Workshop
ICRA 2026

Modeling Kinematics Signatures of Aging via Diffusion and Graph Networks

Yiran Liu[1,2], Eugene Zhao[1,2], Karim Habashy[1,2], Joanna Peng[1,2], Jason Peng[3], Chang Shu[1], Pengcheng Xi[1]

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

Abstract

It is well established that aging significantly shapes human movement — altering gait speed, joint range of motion, and stride consistency in ways that are clinically meaningful but visually subtle. For assistive robots that must anticipate and safely interact with older adults, models that lack age-aware motion priors pose a safety risk. We ask: ***can data-driven models learn genuine age-related biomechanical variation, and generate motions that faithfully reflect those differences?***

We present a three-component framework, evaluated end-to-end on 450 motion clips from 138 subjects (ages 21–86):

• A **data processing pipeline** converting clinical C3D motion-capture recordings into HumanML3D skeletal graph representations, resolving legacy artifacts (axis misalignment, marker mislabeling, mesh inflation) while preserving age-related kinematic cues.

• A **Spatio-Temporal Graph Convolutional Network** (ST-GCN) fine-tuned to classify motion clips into Young, Mid-Age, and Elderly groups, achieving **64.6% validation accuracy** — 31.6 points above the random baseline — and extracting a 32-dimensional latent age embedding z-age that encodes biomechanical gait signatures.

• A **LoRA-adapted Motion Diffusion Model** (MDM) conditioned on discrete age tokens to generate text-driven, age-conditioned motion. Evaluation of 9,000 generated clips shows the model successfully reproduces spatial kinematic constraints.

These findings identify spatiotemporal consistency, conditioning strength, and dataset scale as the critical bottlenecks for age-aware motion synthesis, moltivating the collection of larger, more diverse human motion datasets spanning age groups and health conditions.

Van Criekinge Dataset and Data Proccessing

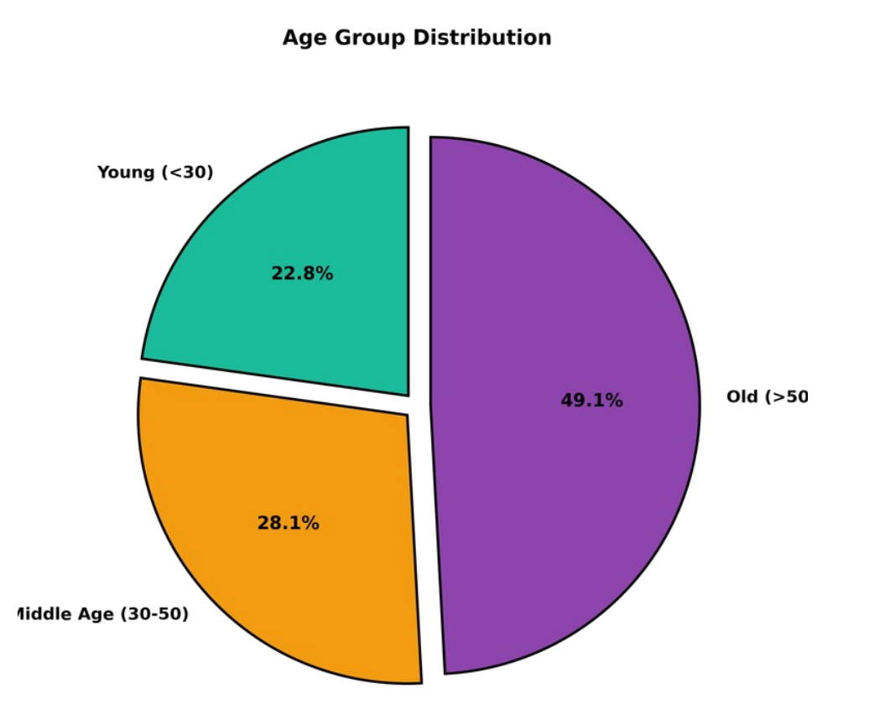
58
Data Points

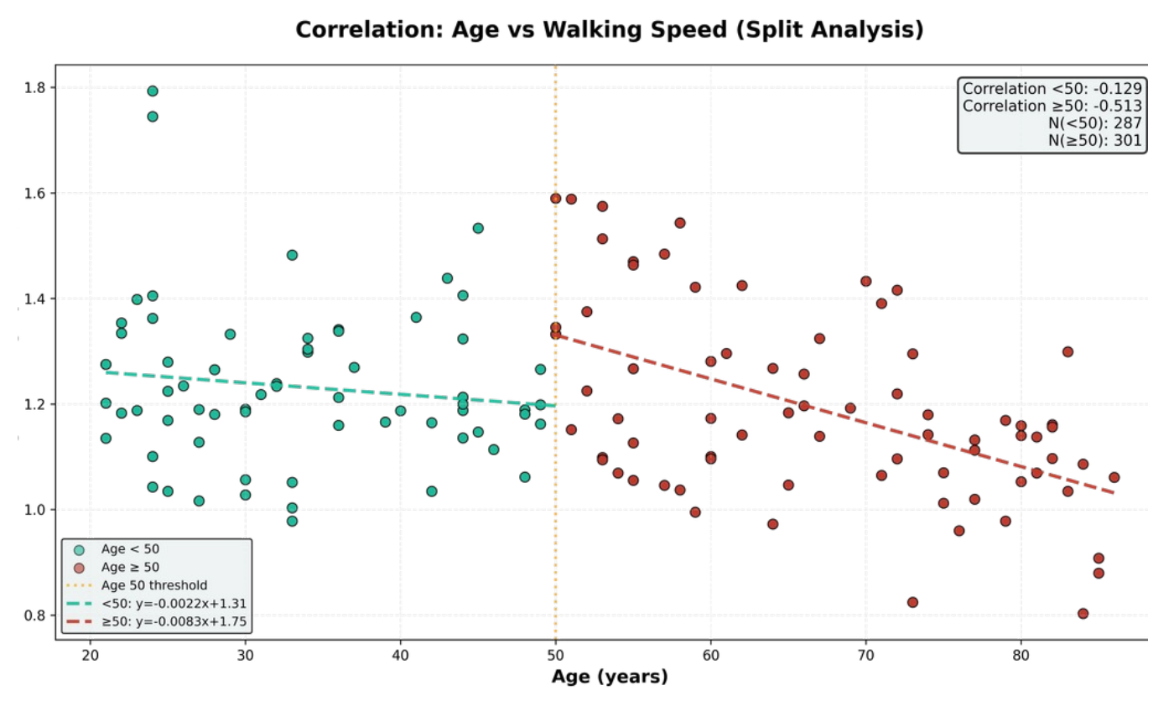
138
Subjects

21–86
Age Range

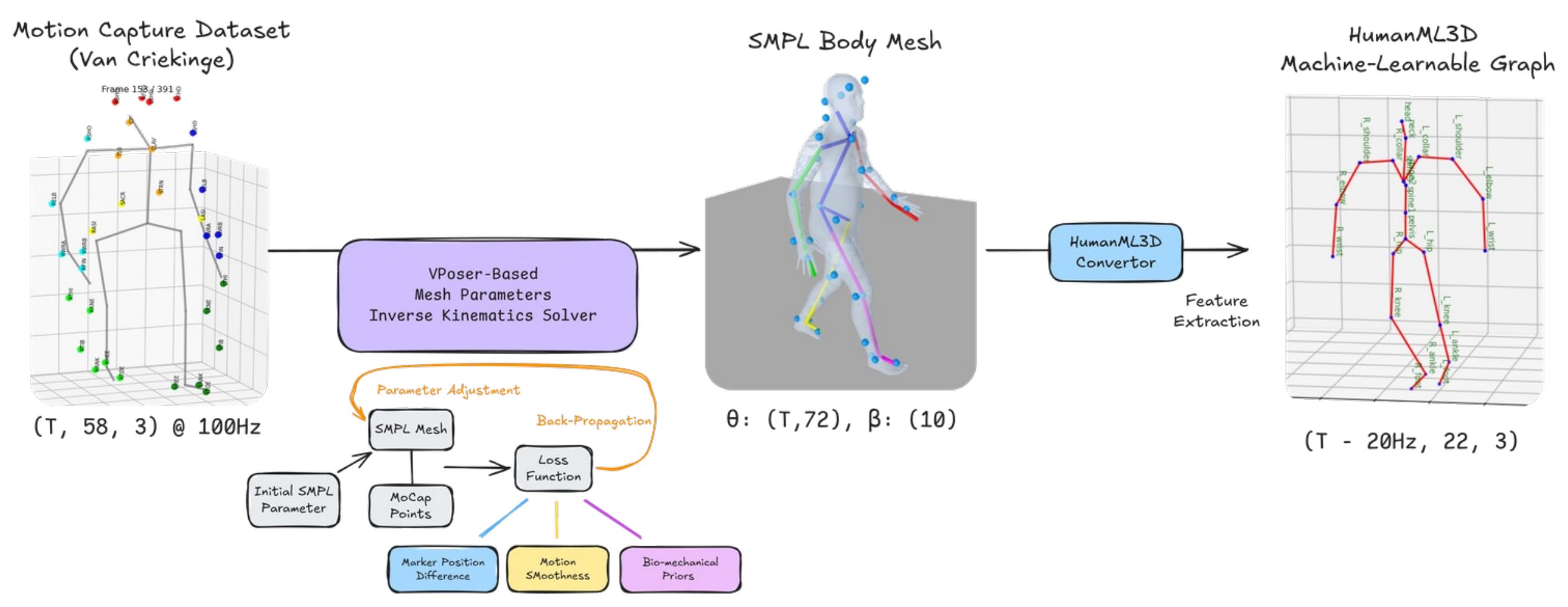
100Hz
Capture

588
Trials





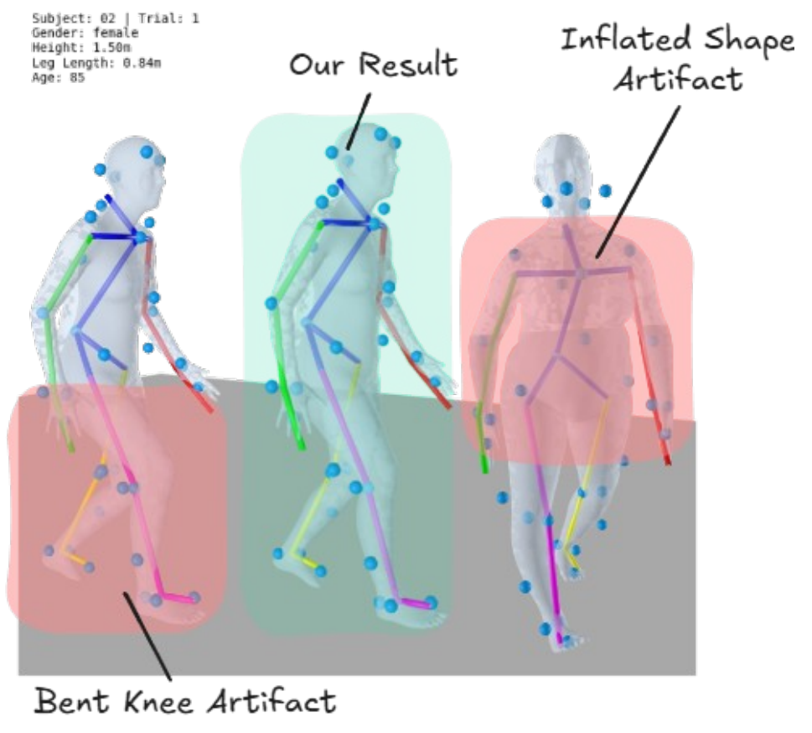
Clinical marker data are converted into a machine-learnable graph format through Vertex-based skin mapping, model fitting and feature extraction and model fitting, with noise removal and temporal smoothing to preserve motion quality.



Mesh Inflation
Over-estimated β params → unrealistic proportions

Crouch Walking
Spurious knee flexion from marker occlusion

✓ **Our Pipeline (Vposer-based)**
Learned pose prior regularizes solution, non-linear optimizer minizes pose difference

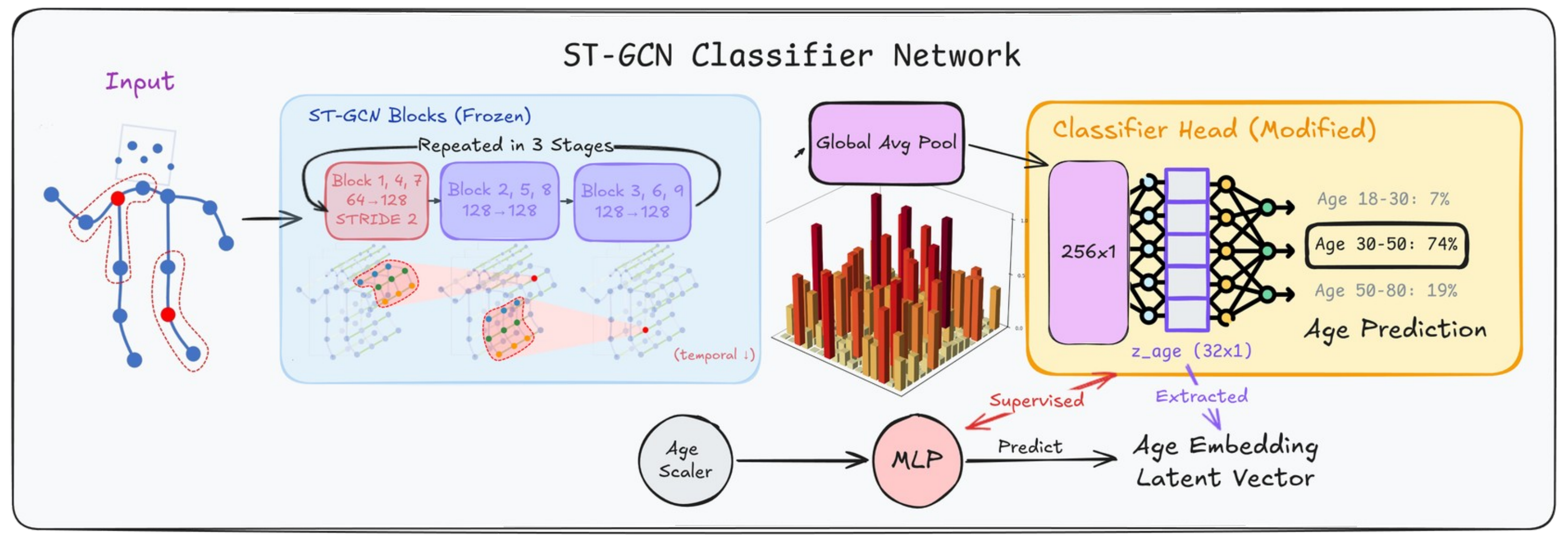

Subject: 02 | Trial: 1
Gender: female
Height: 1.50m
Leg Length: 0.84m
Age: 85

Model Architecture

A **Spatio-Temporal Graph Convolutional Network** trained to discriminate between three age groups: Young (<30), Mid-age (30–60), and Elderly (>60) from skeleton motion.

The backbone is initialized from weights pretrained on NTU RGB+D, while the modified classifier head is retrained on Van Criekinge data.

A compact classification head (256 → 32 → 3) produces age-group predictions; the 32-dimensional penultimate layer serves as **z_age** — a biomechanically-grounded latent encoding of gait age



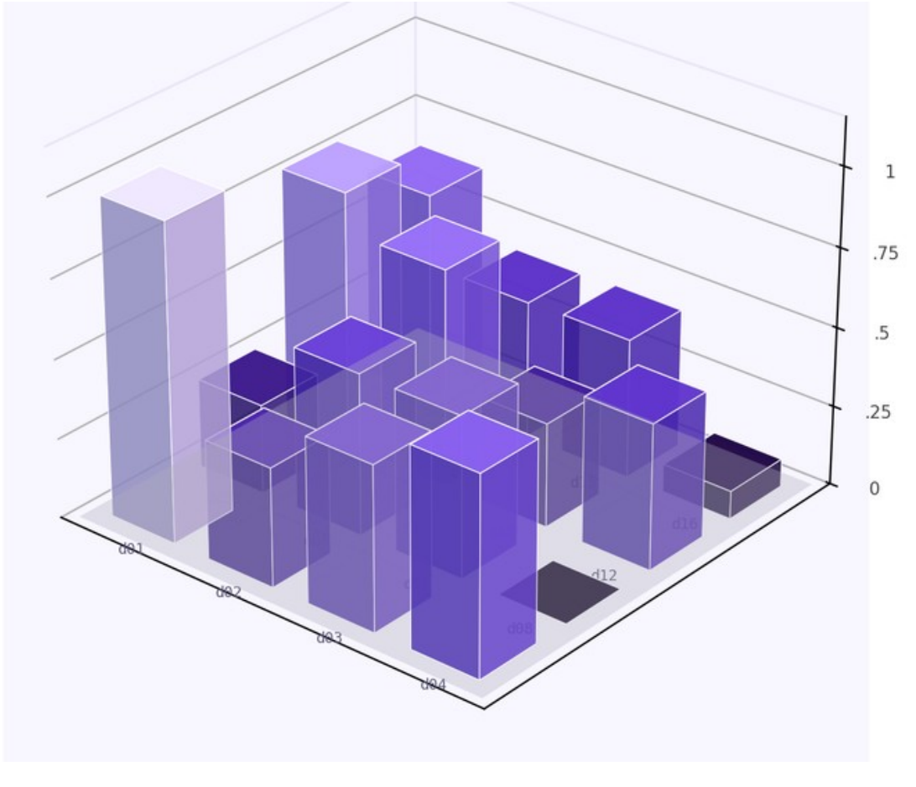
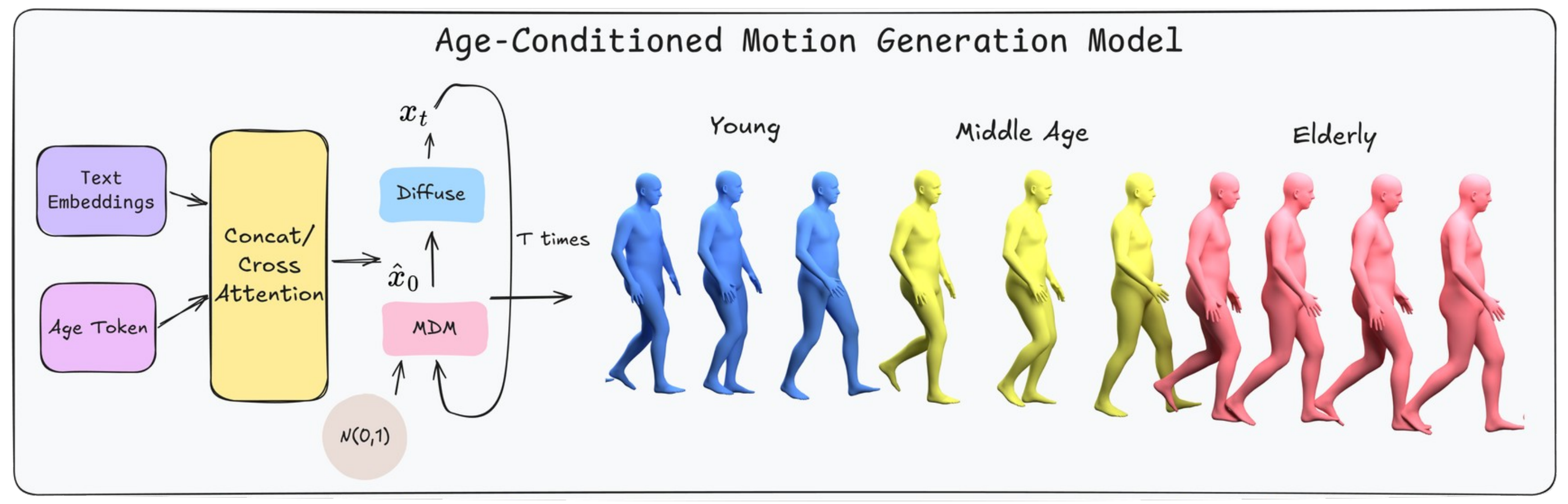


Figure: A single 16-dim latent embedding vector z_age, extract from skeleton motion, visualized as 4x4 grid.



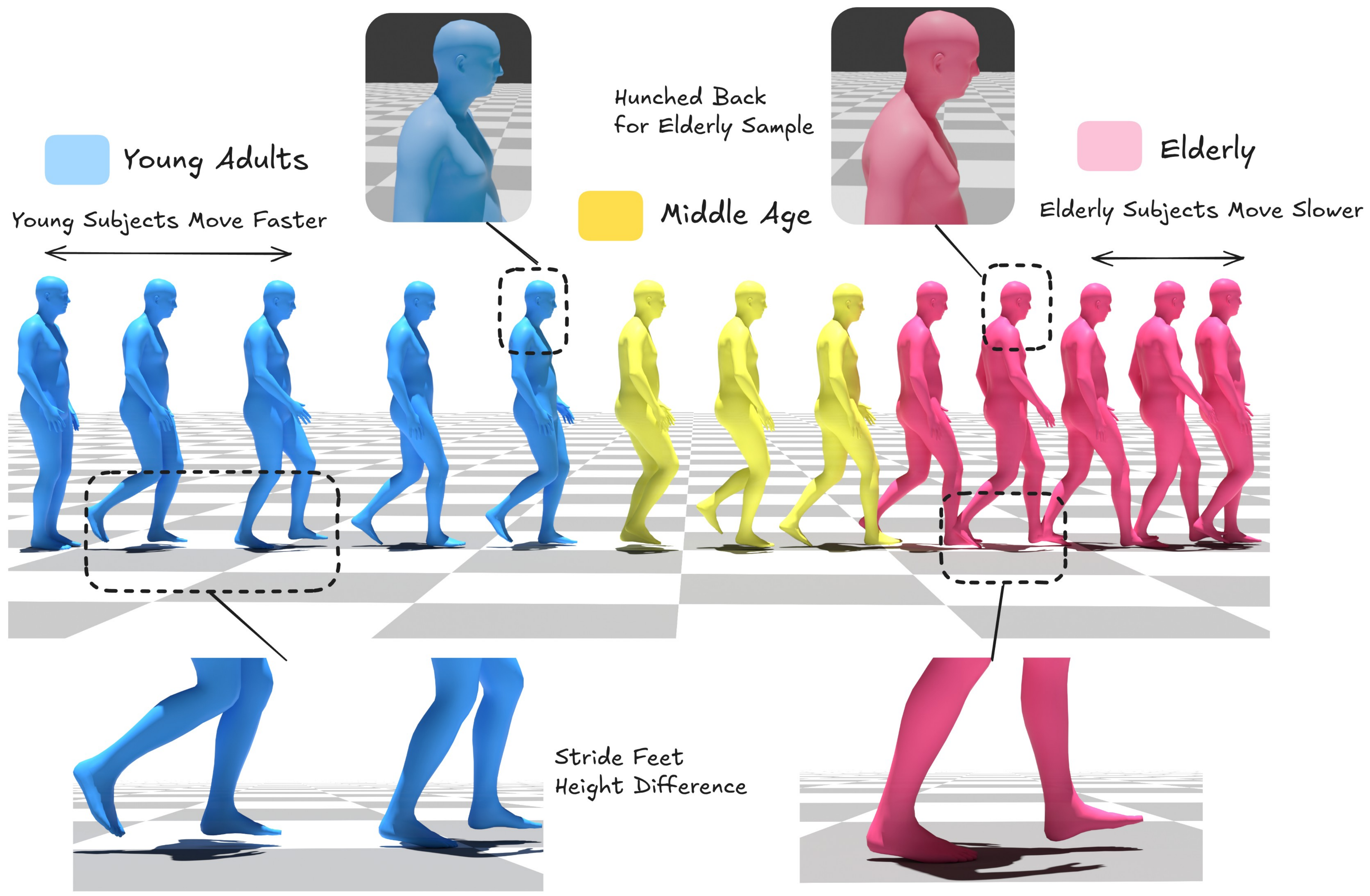
The diffusion model follows the Human Motion Diffusion Model (MDM) architecture. The forward diffusion process is:

$$q(x_t|x_0) = \mathcal{N}\left(x_t; \sqrt{\overline{\alpha}_t}x_0, (1-\overline{\alpha}_t)I\right)$$

where $xt \in \mathbb{R}B \times T \times {}^{263}$ is the noisy motion sequence. Age conditioning is injected by concatenating discrete age token with the text embedding in the cross-attention layers of the denoising transformer. The training objective minimizes the simple noise-prediction loss:

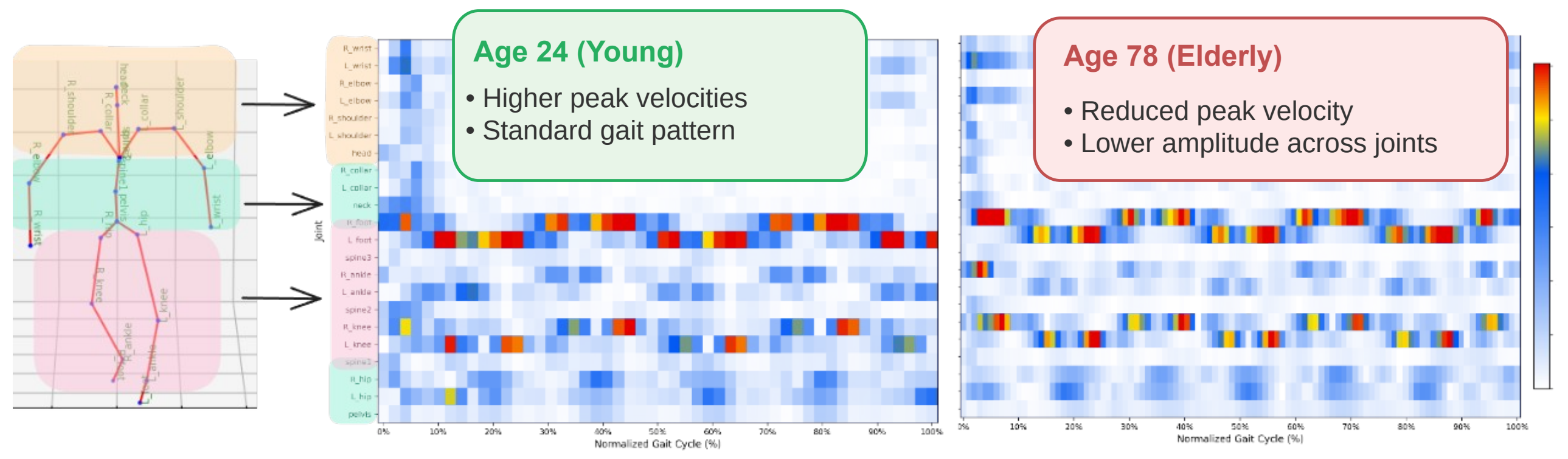
$$\mathcal{L} = \mathbb{E}_{x_0,\varepsilon,t}\left[|\varepsilon - \varepsilon_\theta(x_t, t, z_{\text{text}}, z_{\text{age}})|^2\right]$$

Generation Results



Results Analysis

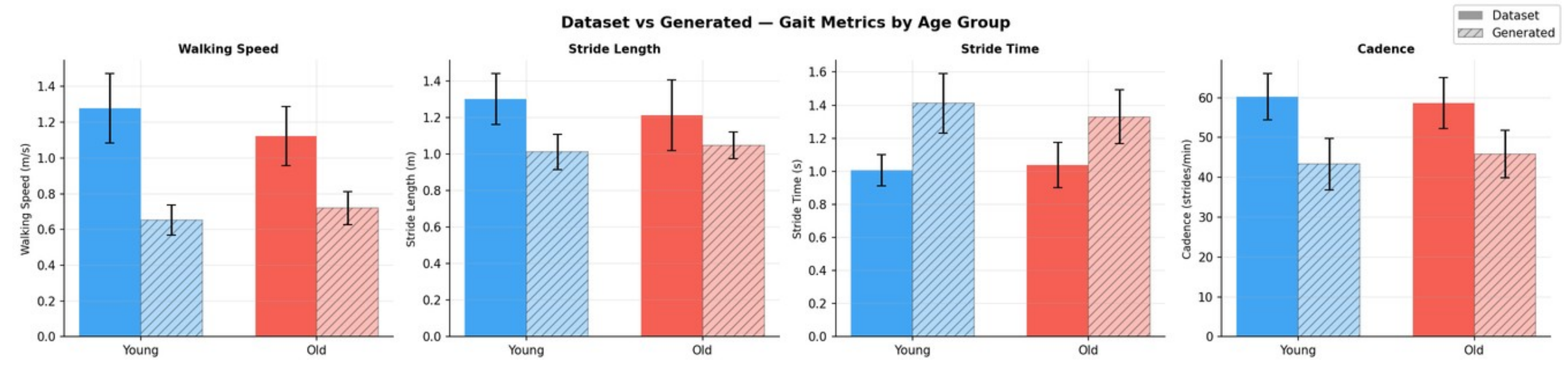
In the dataset, age-discriminative kinematic signals are preserved after conversion — confirming pipeline fidelity. We observe that joint angular velocity (rad/s) decreases consistently as age increase, reflecting biomechanical changes caused by aging.



Age 24 (Young)
- Higher peak velocities
- Standard gait pattern

Age 78 (Elderly)
- Reduced peak velocity
- Lower amplitude across joints

Quantitative analysis of the ground-truth dataset confirms that aging signals are preserved end-to-end through the conversion pipeline: walking speed decreases from 1.22 to 1.10 m/s, stride length from 1.34 to 1.26 m, and Knee Range-of-Motion from 54.9° to 51.5° as subjects age from Young to Elderly. Critically, the LoRA-adapted generative model meaningfully reflects these differences — Elderly-conditioned clips reproduce measurably reduced joint angular velocities across the gait cycle, consistent with the dataset baseline. However, spatiotemporal consistency remains unresolved: generated stride variability uniformly overshoots clinical reference values (16–19% vs. 8.9–15% CV), revealing that diffusion stochasticity currently overwhelms the age conditioning signal for temporal gait regularity.



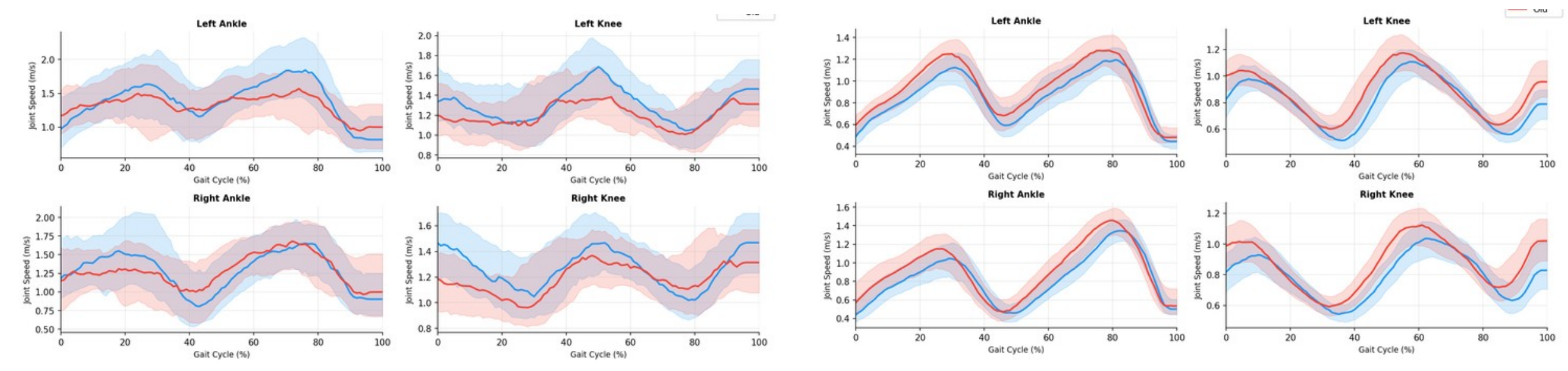


Figure: Distribution of Joint velocity across a normalized gait cycle for **elderly (red)** and **young (blue)** subjects. (Left 4) **Dataset**, (Right4) Generated Samples

Discussion & Future Work

Limitations:

- Small and imbalanced dataset: 450 clips from 138 subjects inevitably leads to ST-GCN overfitting, with Adult mid-range subjects being the primary misclassification failure mode.
- Current LoRA prototype conditions generation on discrete age tokens (Young / Mid / Elderly), which is too coarse to capture gradual, individual-level biomechanical change.
- Diffusion stochasticity overwhelms the conditioning signal for temporal gait regularity, producing stride variability that uniformly overshoots clinical baselines.

Future Directions:

- The natural next step is to replace discrete token conditioning with continuous latent conditioning: the 32-dimensional z_age embedding extracted from the ST-GCN bottleneck layer will be injected directly into the MDM cross-attention layers, enabling generation to be steered by a biomechanically grounded, subject-specific age representation rather than a coarse category label.
- More broadly, this work motivates the collection of larger, higher-quality human motion datasets spanning older adults, varying health conditions, and daily living activities. Such datasets are essential for training demographic-aware motion priors capable of supporting safe and anticipatory human-robot interaction in aging-in-place environments.

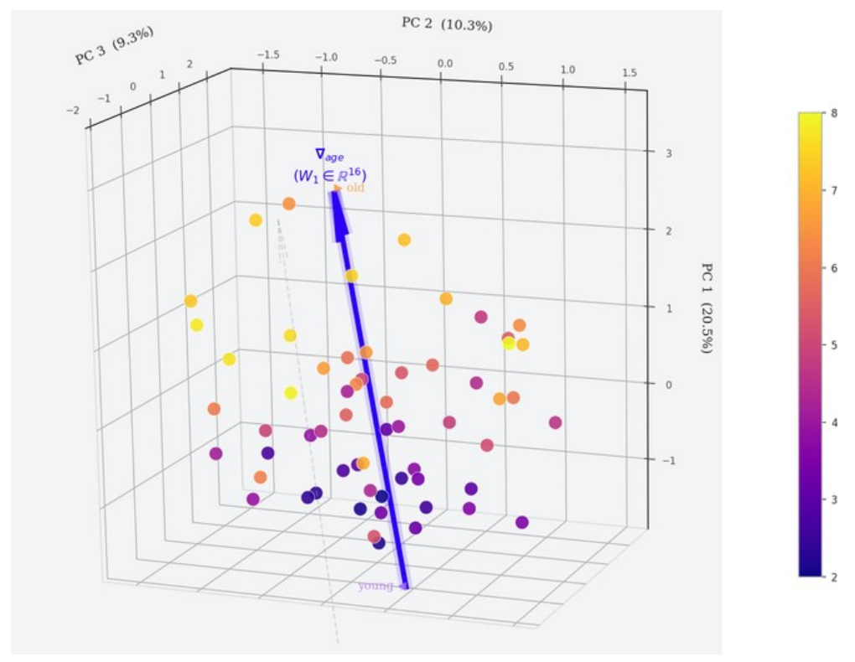


Figure: PCA of z_age embeddings (60 subjects, ages 20–80) reveals moderate-weak age-correlated structure in the current latent space, indicating the classifier has not yet learned age-discriminative representations.
This motivates refinement in model architecture and retraining with larger, more balanced and action-rich dataset.

Acknowledgments

This work was conducted at the National Research Council of Canada (NRC) under the Aging in Place (AIP) challenge program. Compute resources was provided by the ***NRC Trixie HPU cluster****.*

Key references:

[1] Guo et al. — Generating Diverse and Natural 3D Human Motions from Text. CVPR 2022.
Tevet et al. — Human Motion Diffusion Model. ICLR 2023.
[2] Yan et al. — Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition. AAAI 2018
[3] Van Criekinge et al. — A full-body motion capture gait dataset of 138 able-bodied adults across the life span and 50 stroke survivors.