Zero-shot Gait Classification with Diffusion Models

Xiaodong Guan xiaodong.guan.21@ucl.ac.uk
Robert Gray r.gray@ucl.ac.uk
Ashwani Jha ashwani.jha@ucl.ac.uk
Parashkev Nachev p.nachev@ucl.ac.uk
Queen Square Institute of Neurology, University College London

Editors: Under Review for MIDL 2025

Abstract

Movement disorders such as Parkinson's disease are characterised by complex abnormalities of body motion that resist precise, replicable, and scalable quantification. Subjective clinical scores–the established standard–are limited in expressivity and vulnerable to intra-observer variation; wearable sensor-based methods offer objectivity but with limited anatomical sampling. Remote video-based approaches could deliver both highly expressive and objective quantification of motion, but sufficient labelled samples are hard to obtain under clinical data regimes. Here we develop a diffusion model-based, zero-shot, and human-interpretable approach to gait assessment from video-derived pose data and evaluate it in Parkinson's Disease. Capable of detecting subtle changes in body motion without explicit training, it shows potential for an accurate, robust, and scalable solution, addressing the major limitations of existing methods.

Keywords: Gait Analysis, Generative Models, Diffusion Models.

1. Introduction

Movement disorders such as Parkinson's Disease (PD) involve complex abnormalities of body motion that are critical to diagnosis, monitoring, and treatment selection. Gait is a key aspect here, since it is both frequently affected and of great functional significance. Clinicians typically use subjective scoring systems such as Performance-Oriented Mobility Assessment(Tinetti et al., 1986) or Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS)(Goetz et al., 2008), that rely on observational judgments. However, their subjective nature poses challenges in reproducibility and precise quantification.

Emerging machine-learning approaches quantify gait parameters through computer-vision methods, analyzing features such as feet distance(Verlekar et al., 2018), swing velocity(Eltoukhy et al., 2017), 2D body pose(Rupprechter et al., 2021; Jinila et al., 2022; Tan et al., 2024; Nõmm et al., 2016), gait energy images(Ortells et al., 2018), and body keypoints in Cartesian space(Kaur et al., 2022); or through wearable-device-based feature extraction(Han et al., 2023; Moreau et al., 2023). Although promisingly performant on selected test datasets, these methods have limitations such as reliance on specially designed environments, customized camera setups, tailored sensors, and abundant labelled data. Moreover, most measurements are conducted in Cartesian or 2D space, lacking translation- and appearance-invariance, thus contributing to poor generalizability.

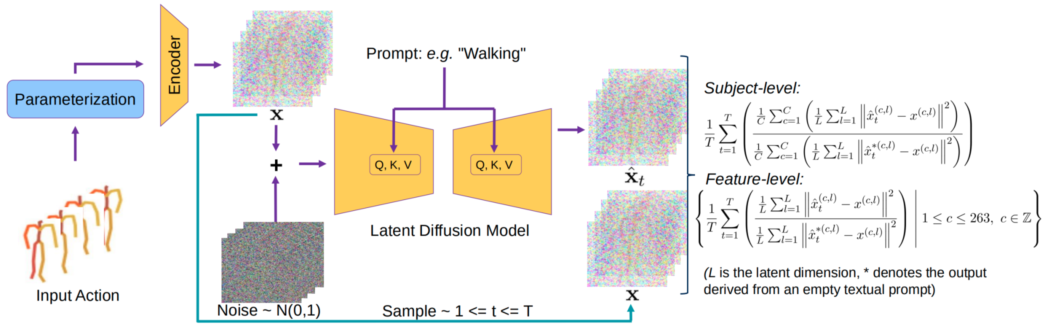

Figure 1: Proposed framework for quantifying the deviation of a test action from its prediction under a given prompt, yielding a measure of movement anomaly.

To address these limitations, here we propose a zero-shot diffusion model pretrained on diverse human motions, quantifying the deviation of input actions from specific textual action descriptions. The approach offers a high-fidelity mapping between textual descriptions and 3D pose sequences, allowing for scale-, translation-, and viewpoint-invariant gait representations. At test time, the model evaluates gait clips against generated motions, producing a robust anomaly score that aligns well with clinician assessments, without requiring task-specific training or annotated datasets. Evaluated in PD, this approach enables sensitive, scalable, and automated gait assessment for early disease detection in clinical and real-world settings, and is transferable to other clinical scenarios.

2. Method

We parameterized body movements over 1-1.6 second intervals, indexed by 24 video-derived joint positions sampled at 25Hz, using axis-angle representation, and converted them into the HumanML3D format (Guo et al., 2022), neutralizing the effects of body shape. A human movement latent diffusion model (Tevet et al., 2022) was used to model the distribution of movements, conditioned on their textual descriptions, providing an index of the anomaly of test movements qualified under a normative description(Li et al., 2023). Instead of evaluating predicted noise(Li et al., 2023), we compared predicted vs original latent features at each sampled time-step, formulating a subject-level anomaly measure as the mean squared error across all latent dimensions and feature channels, and a feature-level anomaly measure as the mean squared error across all latent dimensions. Both measures are normalized using values derived from an empty textual prompt, then averaged over time (see Figure 1).

3. Experiments and Discussion

Result We evaluated the proposed framework on gait data from 62 patients, divided into 4 groups based on MDS-UPDRS Part 3 gait scores assessed within normal medical treatment regime (0–3 in this cohort, where 0 indicates normal gait and higher scores reflect progressively worsening motor function), using 400 gait clips (100 per group) recorded with various devices, including smartphones and webcams. The evaluation focused on the correspondence between our anomaly measure and MDS-UPDRS scores. For both subject- and feature-level measures, we used the first 900 of the 1000 total time-steps, as they provided the greatest distinction between prompts. Using the prompt ”Walking”, the subject-level measure yielded a Spearman correlation coefficient of 0.7011 (p <1e-5). A one-way ANOVA comparing between- and within-group variance produced an F-value of 186.14 (p <1e-5). The threshold of significance for both tests was set at 0.05. The parameterized motion data consisted of 263 dimensions, capturing interpretable features including Cartesian joint positions, joint rotations (as 6D matrices(Zhou et al., 2019)), and linear velocities, for which feature-level anomaly measures are derived. In Figure 2, the first three panels show comparisons of each non-zero MDS-UPDRS score level against normal (MDS-UPDRS=0) gait, showing that the gaits of MDS-UPDRS > 0 groups exhibit greater errors than those from the MDS-UPDRS=0 group, with concentrations in the lower limbs and distal joints such as the feet and hands. Joint position and rotation errors were also prominently affected in the head and knee regions. Velocity-based errors were larger at the hip and spine, likely due to rigidity causing greater deviation from typical gait patterns. The right-most panel shows the correlation between feature-level anomaly measures and MDS-UPDRS scores, further highlighting that shoulder, lower limb, and head regions are most strongly associated with MDS-UPDRS when measuring Cartesian position; distal joints (hands, head, feet) dominate for rotations; and the pelvic region contributes most in the velocity domain. These results suggest that reduced dexterity, flexibility, and impaired joint coordination are key contributors to gait abnormalities in PD. The anatomical distribution of error also indicates imbalanced movement, consistent with PD-related asymmetry.

Conclusion We introduce an interpretable diffusion model-based gait classification framework for quantifying abnormalities of gait that combines expressivity with objectivity, and is operable within the few-label data regimes clinical scenarios impose. In the context of PD, our zero-shot approach enables the quantification of abnormalities–both feature- and subject-level–with good fidelity without explicit training, facilitating implementation, and broadening access to automated analysis. The high correlation with MDS-UPDRS scores–obtained zero-shot–leaves room for substantial benefit from fine-tuning as larger volumes of data become available. Conditioning on highly expressive text labels enables ready extension of the framework to richly defined classification tasks, with utility across all clinical scenarios, across movement disorders and beyond, where quantification of body motion is critical to clinical management.

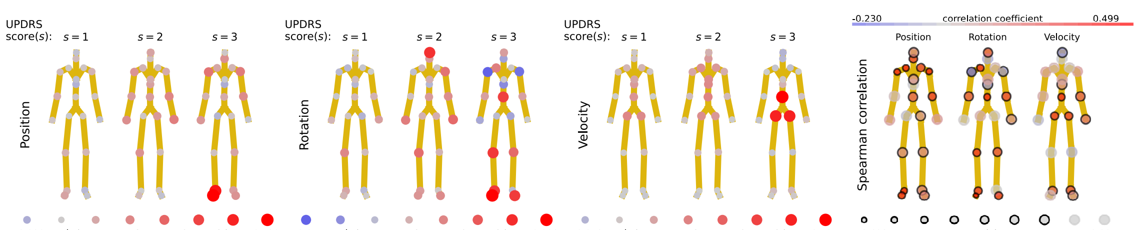

Figure 2: Anatomical projection of joint-wise error magnitudes across groups, along with the correlation strength contributed by errors at each joint. Outlined circles indicate regions below the corrected statistical significance threshold (p = 0.0166), with correlations computed independently for each data component.

References

Moataz Eltoukhy, Christopher Kuenze, Jeonghoon Oh, Marco Jacopetti, Savannah Wooten, and Joseph Signorile. Microsoft kinect can distinguish differences in over-ground gait between older persons with and without parkinson's disease. *Medical engineering & physics*, 44:1–7, 2017.

Christopher G Goetz, Barbara C Tilley, Stephanie R Shaftman, Glenn T Stebbins, Stanley Fahn, Pablo Martinez-Martin, Werner Poewe, Cristina Sampaio, Matthew B Stern, Richard Dodel, et al. Movement disorder society-sponsored revision of the unified parkinson's disease rating scale (mds-updrs): scale presentation and clinimetric testing results. *Movement disorders: official journal of the Movement Disorder Society*, 23(15):2129–2170, 2008.

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 (CVPR)*, pages 5152–5161, June 2022.

Yi Han, Xiangzhi Liu, Ning Zhang, Xiufeng Zhang, Bin Zhang, Shuoyu Wang, Tao Liu, and Jingang Yi. Automatic assessments of parkinsonian gait with wearable sensors for human assistive systems. *Sensors*, 23(4):2104, 2023.

Bevish Jinila et al. Vision-based gait analysis for real-time parkinson disease identification and diagnosis system. *Health Systems*, 13(1):62, 2022.

Rachneet Kaur, Robert W Motl, Richard Sowers, and Manuel E Hernandez. A vision-based framework for predicting multiple sclerosis and parkinson's disease gait dysfunctions—a deep learning approach. *IEEE journal of biomedical and health informatics*, 27(1):190–201, 2022.

Alexander C Li, Mihir Prabhudesai, Shivam Duggal, Ellis Brown, and Deepak Pathak. Your diffusion model is secretly a zero-shot classifier. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2206–2217, 2023.

Caroline Moreau, Tiphaine Rouaud, David Grabli, Isabelle Benatru, Philippe Remy, Ana-Raquel Marques, Sophie Drapier, Louise-Laure Mariani, Emmanuel Roze, David Devos, et al. Overview on wearable sensors for the management of parkinson's disease. *npj Parkinson's Disease*, 9(1):153, 2023.

Sven Nõmm, Aaro Toomela, Martti Vaske, Dan Uvarov, and Pille Taba. An alternative approach to distinguish movements of parkinson disease patients. *IFAC-PapersOnLine*, 49(19):272–276, 2016.

Javier Ortells, María Trinidad Herrero-Ezquerro, and Ramón A Mollineda. Vision-based gait impairment analysis for aided diagnosis. *Medical & biological engineering & computing*, 56:1553–1564, 2018.

Samuel Rupprechter, Gareth Morinan, Yuwei Peng, Thomas Foltynie, Krista Sibley, Rimona S Weil, Louise-Ann Leyland, Fahd Baig, Francesca Morgante, Ro'ee Gilron, et al. A clinically interpretable computer-vision based method for quantifying gait in parkinson's disease. *Sensors*, 21(16):5437, 2021.

Vincent Wei Sheng Tan, Wei Xiang Ooi, Yi Fan Chan, Tee Connie, and Michael Kah Ong Goh. Vision-based gait analysis for neurodegenerative disorders detection. *Journal of Informatics and Web Engineering*, 3(1):136–154, 2024.

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.

Mary E Tinetti, T Franklin Williams, and Raymond Mayewski. Fall risk index for elderly patients based on number of chronic disabilities. *The American journal of medicine*, 80 (3):429–434, 1986.

Tanmay T Verlekar, Luís D Soares, and Paulo L Correia. Automatic classification of gait impairments using a markerless 2d video-based system. *Sensors*, 18(9):2743, 2018.

Yi Zhou, Connelly Barnes, Jingwan Lu, Jimei Yang, and Hao Li. On the continuity of rotation representations in neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5745–5753, 2019.