Diffusion-based Personalized Pathology Disentanglement for Impaired Gait Analysis

Xiaoyue Wan, Xu Zhao*

School of Automation and Intelligent Sensing, Shanghai Jiao Tong University, China
{sherrywaan, zhaoxu}@sjtu.edu.cn

Abstract

In the context of global population aging, the prevalence of neurodegenerative diseases is rapidly increasing. Vision-based impaired gait analysis emerges as a promising alternative for automatic and non-invasive diagnosis. While prior efforts have advanced either accuracy or interpretability of gait analysis, few have effectively addressed both aspects in a unified framework. To bridge this gap, we propose DPPD, a Diffusion-based Personalized Pathology Disentanglement model that jointly performs quantitative gait scoring, dementia subtyping, and qualitative anomaly highlighting. Motivated by the observation that pathological gait features exhibit stronger inter-class separability across different gait severity than raw features, DPPD is proposed based on the subject-specific pathology disentanglement perspective. Specifically, it comprises three key components: (1) a 3DmotionBERT for encoding gait representation from 3D human pose sequences estimated, (2) a latent diffusion-based Gait Denoiser for generating personalized normal gait features, and (3) a Dual Pathology Disentanglement mechanism that captures both static pose and dynamic motion pathological representation from the residual between raw and normal gait features. These disentangled pathologies further enable quantitative classification and qualitative anomaly highlighting. Experiments on the PDGait and 3DGait datasets demonstrate that DPPD outperforms state-of-the-art methods in classification accuracy while providing reliable and interpretable visualizations of gait anomalies.

Code — https://github.com/sherrywan/DPPD

Introduction

With the global population aging, the prevalence of neurodegenerative diseases (NDDs), such as Parkinson's disease (PD), Alzheimer's disease (AD), and Dementia with Lewy Bodies (DLB), is rising steadily (Jiang et al. 2025). The pressing need for efficient diagnostic techniques has brought intelligent gait analysis into growing focus, as gait impairment is one of the most prominent symptoms (Verghese et al. 2002). Previous studies have attempted various sensing modalities to capture gait data and conduct impaired gait analysis, including gait scoring (e.g., rating gait anomalies on a 0–4 scale) and dementia subtyping (e.g., distinguishing between PD, AD, and DLB). Among these, vision-based methods are gaining popularity due to their cost-effectiveness, non-invasiveness, and suitability for both clinical and remote assessments.

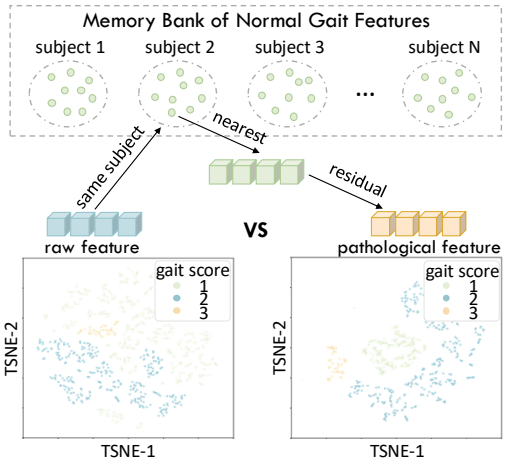

Figure 1: A t-SNE-based comparative analysis of raw and pathological gait features. Pathological features are computed as the residuals between raw features and their nearest normal counterparts within the same subject, retrieved from a pre-built memory bank of normal gait features.

To enable real-world clinical adoption, video-based gait analysis models must not only achieve high accuracy but also provide transparent and interpretable evidence of decision-making. However, balancing these two objectives remains a challenge. Early approaches have made considerable efforts to improve the accuracy of gait analysis tasks based on videos or estimated 3D human poses. Various architectures are adopted to enhance gait temporal modeling, such as Long Short Term Memory networks (LSTM) (Albuquerque et al. 2021), Spatiotemporal Graph Convolution networks (ST-GCNs) (Sabo et al. 2022), and Transformers (Wang et al. 2023; Adeli et al. 2024). Vision-language models (VLMs) are also introduced to mitigate data scarcity, using text encoders to improve representation stability (Wang et al. 2024). Despite achieving competitive performance, they typically operate as a "black box", lacking intuitive in-

*Corresponding author.

terpretability. Nowadays, with the advent of large language models (LLMs), new opportunities emerge for generating diagnostic rationales (Yeh et al. 2024). Notably, Wang et al. propose combining gait scoring with explainable reasoning using LLM-based chain-of-thought (Wang, Bobenrieth, and Seo 2025). Despite these advances, since the existing clinical gait data is insufficient to effectively train LLMs, the improvement of its interpretability seems to come at the expense of accuracy. In this work, we aim to bridge this gap by developing a model that ensures accuracy while providing interpretability for impaired gait analysis with 3D human poses as input, striving to be closer to practical application in clinical auxiliary diagnosis.

Clinicians typically assess gait severity by identifying pathological symptoms, guided by the Movement Disorder Society – Unified Parkinson's Disease Rating Scale (MDS-UPDRS) (Goetz et al. 2008). These symptoms are commonly recognized as deviations from normal gait patterns (Wolfson et al. 1990). However, most existing models overlook this pathological perspective, instead directly performing gait scoring or dementia subtyping on raw gait representations. To investigate the value of explicitly pathology modeling, we conduct a comparative analysis on PDGait dataset (Shida et al. 2023). Specifically, using 3DmotionBERT (see Method) to encode 3D pose sequences, we visualize both original and pathological features via t-SNE (Maaten and Hinton 2008). Here, pathological deviation is defined as the residual between a given gait representation and its nearest normal counterpart within the same subject. As shown in Fig. 1, residual features exhibit better inter-class separability across severity levels, suggesting the potential of explicitly disentangling pathology to benefit gait analysis. Moreover, this disentanglement enables the localization of anomalies with obvious pathology and further provides intuitive visual evidence to support quantitative decision-making, offering interpretability for model inference.

The key challenge, however, lies in obtaining the normal gait representation of each subject, which is often unavailable in real-world or even a dataset. To address this, we draw inspiration from recent advancements in diffusion-based anomaly detection (He et al. 2024; Hu et al. 2024; Zhang et al. 2025b), and propose to generate a subject-specific normal gait representation through a diffusion model trained only on normal data. The residual between the original and generated gait features then serves as a representation of pathological deviation. Building on this insight, we propose a novel impaired gait analysis framework, the Diffusion-based Personalized Pathology Disentanglement model, named DPPD.

DPPD consists of three key modules. First, a 3Dmoiont-BERT encoder is pre-trained with a masked 3D pose sequence reconstruction task in an unsupervised manner, enabling frame-joint-level feature extraction from 3D pose sequences and supporting fine-grained anomaly localization. A temporal consistency loss is introduced to ensure that the residual feature distance correlates with actual pose differences, thereby reflecting anomaly severity as well. Second, a latent diffusion-based Gait Denoiser (GD) is employed to reconstruct normal gait features. To achieve personality modeling, we introduce a first-frame feature condition and a subject-level personality contrastive loss. Finally, a Dual Pathology Disentanglement (DPD) module is proposed to extract both pose-level and motion-level pathological deviations. These disentangled features are then utilized for downstream tasks, including quantitative gait scoring, dementia subtyping and qualitative anomaly highlighting. Our contributions can be summarized as follows:

- We propose a novel impaired gait analysis framework, DPPD, which explicitly disentangles pathological features for the first time, enabling both accurate and interpretable gait assessment for NDDs.
- We introduce a novel module, DPD, covering pathologies from static to dynamic, providing interpretable visualizations aligned with specific anomaly categories.
- Extensive experiments demonstrate that our DPPD outperforms existing methods in both gait scoring and dementia subtyping, while also providing promising interpretability through visualization.

Related Work

Video-based Gait Analysis in NDDs. Previous studies have employed various sensors, including force plates (Trosch et al. 1998; Conroy 2008), inertial measurement units (IMUs) (Hsu et al. 2018; Li et al. 2020), and videos (Sarapata et al. 2023), to capture gait data and conduct impaired gait analysis. Among these, vision-based methods are gaining popularity due to their cost-efficiency, non-intrusiveness. Vision-based approaches for impaired gait analysis can be broadly categorized into heuristic and data-driven methods. Heuristic methods typically rely on handcrafted features with explicit physical semantics, such as stride length (Chee et al. 2009), walking speed (Sabo et al. 2020), and shoulder tilt (Deng et al. 2024). These features are extracted from 3D human poses estimated from video, often guided by clinicians. While offering interpretability, their generalization is limited. In contrast, data-driven approaches leverage deep learning models to learn gait representations directly from videos or 3D pose sequences, including LSTM networks (Albuquerque et al. 2021), ST-GCNs (Sabo et al. 2022), temporal convolutional networks (Zeng et al. 2023), Transformers (Wang et al. 2023; Adeli et al. 2024), and VLMs (Wang et al. 2024). Despite achieving competitive performance, these methods often lack interpretability, which hinders clinical adoption due to limited transparency. To improve interpretability, recent work has explored large language models (LLMs) to enable chain-of-thought reasoning for gait scoring and explanation (Wang, Bobenrieth, and Seo 2025). However, limited availability of clinical gait datasets constrains the effectiveness of LLM-based reasoning, resulting in a trade-off between interpretability and accuracy.

Diffusion-based Anomaly Detection. Diffusion models, notably DDPM (Ho, Jain, and Abbeel 2020), have recently shown great promise in anomaly detection. Trained solely on normal data, these models tend to reconstruct anomalous inputs as their closest normal counterparts. The discrepancy between the input and its reconstruction serves as an

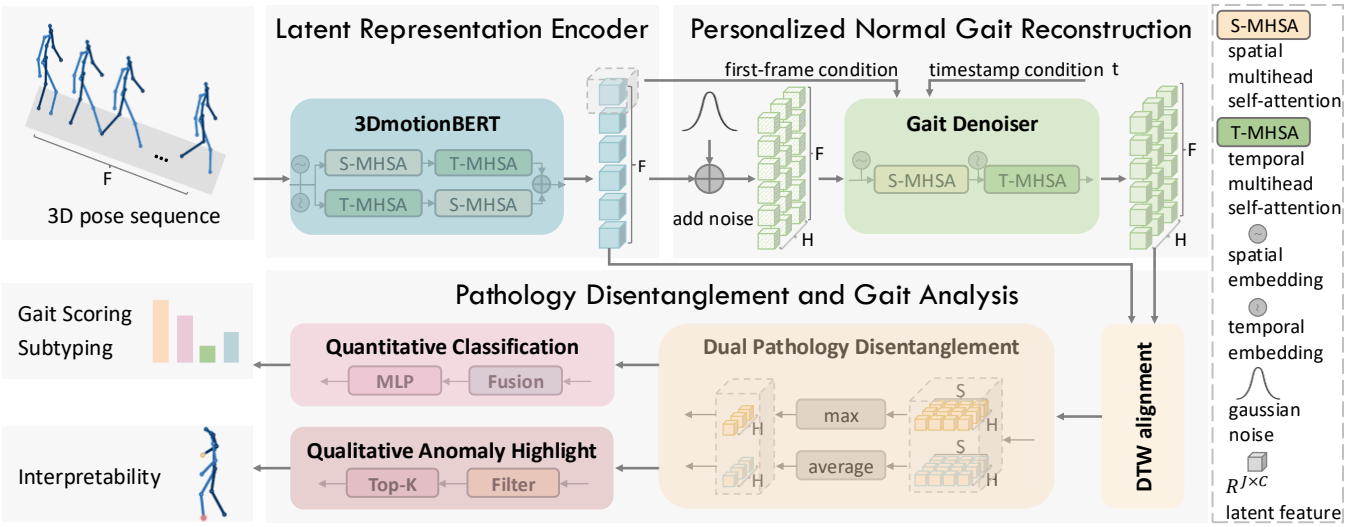

Figure 2: **The framework of DPPD.** A 3D pose sequence is first encoded into a latent representation using 3DmotionBERT. The GD module then generates the corresponding normal gaits from a noisy version of the encoded input. Based on the original and generated normal representations, dual pathological features are extracted through DPD and used for both quantitative gait classification and qualitative anomaly highlighting.

effective signal for localizing anomalies (Liu et al. 2025). For instance, DiffusionAD (Zhang et al. 2025a) proposes a norm-guided one-step denoising scheme to accelerate inference. CLIPrompt (Liu et al. 2024) introduces contrastively learned prompts to guide the diffusion model in detecting domain-specific anomalies. To handle real-world complexity, recent methods extend diffusion models to multi-class and zero-shot settings. DiAD (He et al. 2024) integrates semantic feature guidance to preserve category semantics and spatial structures. DZAD (Zhang et al. 2025b) leverages intermediate denoising steps for zero-shot anomaly detection without external prompts. These advances motivate us to generate normal gait representation via diffusion models.

Method

The inference framework of DPPD is illustrated in Fig. 2. Given a 3D human skeleton pose sequence $x \in \mathbb{R}^{F\times J\times 3}$ extracted from a video with F frames, we first encode it into the latent representation $z \in \mathbb{R}^{F\times J\times C}$ using a pre-trained 3DmotionBERT, where J denotes the number of joints and C is the latent feature dimension. Gaussian noise is then added to z, and the GD module denoises it to generate H subject-specific normal gait representation hypotheses, conditioned on the first-frame feature $z_{f=0} \in \mathbb{R}^{J\times C}$, where f is the frame index. Finally, the DPD module disentangles dual pathological representations from each temporally aligned pair of raw and generated latent gait features. The resulting pathological representations are further used for quantitative gait analysis tasks: gait scoring and dementia subtyping. Meanwhile, it also enables qualitative abnormal highlighting to visualize the underlying pathological evidence for decision-making. All three modules are trained independently, with the others frozen. We detail the architecture and training strategy below.

Latent Representation Encoder

This module aims to encode the 3D pose sequence x into a fine-grained latent gait representation, similar to the first stage of Stable Diffusion (Rombach et al. 2022). Reliable gait analysis for clinicians should determine when and at which joint an abnormality occurs. Existing motion encoders (Tevet et al. 2022; Jiang et al. 2023; Guo et al. 2024) are generally limited to pose or sequence level representation, which are not sufficiently fine-grained for our purpose. While MotionBERT (Zhu et al. 2023) outputs frame-joint-level representations, it is designed with 2D pose as input. Hence, we develop 3DmotionBERT by adapting it to 3D pose sequences. For brevity, we omit architectural details as it is the same as MotionBERT. Specifically, 3DmotionBERT is trained via a masked reconstruction task: the input sequence x is randomly masked at the frame-joint level, then encoded into latent representation $z \in \mathbb{R}^{F\times J\times C}$, which is subsequently reconstructed to $\hat{x}$ through a regression head. The primary training objective is the 3D reconstruction loss:

$$\mathcal{L}_{3D} = \sum_{f=1}^{F}\sum_{j=1}^{J} \| \hat{x}_{f,j} - x_{f,j} \|_2, \tag{1}$$

where $\hat{x}$ is the regressed 3D pose sequence. To further enhance the temporal coherence of the latent representation, we introduce a temporal consistency loss:

$$\mathcal{L}_{tem} = \sum_{f=2}^{F}\sum_{j=1}^{J} \| z_{f,j} - z_{f-1,j} \|_2, \tag{2}$$

which encourages adjacent frames to have consistent representations. This term offers two advantages: (1) it promotes a smoother latent motion trajectory, which benefits downstream diffusion modeling, and (2) it aligns latent-space distances with actual pose similarities, thereby enhancing the

semantic consistency between residual representations and pathological severity. The final loss combines both terms:

$$\mathcal{L} = \mathcal{L}_{3D} + \mathcal{L}_{tem}, \tag{3}$$

with equal weighting, as they are empirically observed to have similar magnitudes.

Personalized Normal Gait Reconstruction

To reconstruct the normal gait representation of the patient being assessed, we propose a latent diffusion model, referred to as GD, which learns the distribution of normal gait representations in the latent space. It adopts the MixSTE architecture (Zhang et al. 2022) and is trained solely on normal gait samples to capture the distribution of healthy gait representation through a forward diffusion process followed by a reverse denoising process under a Markov chain with T steps. Given the latent representation $n_0 \in \mathbb{R}^{F\times J\times C}$ of a normal gait sequence encoded by 3DmotionBERT, we add Gaussian noise according to the forward diffusion process:

$$n_t = \sqrt{\bar{\alpha}_t} n_0 + \sqrt{1-\bar{\alpha}_t}\epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \mathbf{I}), \tag{4}$$

where n_t is the disturbed representation at diffusion timestamp $t \in [1, T]$, and α_t follows a predefined noise schedule in DDPM (Ho, Jain, and Abbeel 2020). n_t is then fed to GD D_θ, which aims to recover the normal gait representation:

$$\mathcal{L}_\theta = \mathbb{E}_{t,n_0,\epsilon_t} \parallel n_0 - D_\theta(n_t, t) \parallel_2^2 . \tag{5}$$

During inference, each gait representation z is disturbed under timestamp $t = T$, obtaining n_T, and then performs iterative denoising through GD to generate the normal representation $\hat{n}_0$. DDIM sampling (Song, Meng, and Ermon 2020) is applied for I denoising iterations. To account for stochasticity, we generate H reconstruction hypotheses per sequence.

As observed, normal gait varies across individuals due to differences in skeleton and even motion habits. Therefore, it is crucial for GD to incorporate subject-specific characteristics to generate personalized normal gait representations. To address this, we introduce two complementary techniques:

First-frame Condition. To inject personality perception into GD, denoising is conditioned on the first-frame feature $z_{f=0}$ clipped from the original gait representation. Specifically, we add temporal embeddings to $z_{f=0}$ and n_T, concatenate them, and then project back to dimension C using a linear layer. The denoising objective is updated as:

$$\mathcal{L}_\theta = \mathbb{E}_{t,n_0,\epsilon_t} \parallel n_0 - D_\theta(n_t, t, z_{f=0}) \parallel_2^2 . \tag{6}$$

Personality Contrastive Loss. We further introduce a personality contrastive loss to facilitate subject-level representation alignment. This loss encourages the denoised representation to remain close to the raw representations from the same subject, while being distinct from those of other subjects within a training batch:

$$\mathcal{L}_{per} = \frac{1}{N}\sum_{n=1}^{N}\mathcal{L}_n, \tag{7}$$

$$\mathcal{L}_n = -\log \frac{\sum_{m\in\mathcal{P}_n}\exp(\text{sim}(n,m))}{\sum_{k=1}^{N}\exp(\text{sim}(n,k))}, \tag{8}$$

where $\text{sim}(n,m) = \frac{z_n^\top z_m}{\tau}$ denotes the similarity between the gait representations of subjects n and m, scaled by a temperature parameter $\tau = 0.07$. $\mathcal{P}_n = \{m \mid id_m = id_n\}$ is the set of the same subject within this batch, and N is the batch size. The final denoising object is formulated as:

$$\mathcal{L} = \mathcal{L}_\theta + \lambda_{per}\mathcal{L}_{per}, \tag{9}$$

where $\lambda_{per} = 0.01$ is empirically set to ensure that GD can effectively recover the normal gait representation.

Pathology Disentanglement and Gait Analysis

To interpret the learned gait representation, we regress it back to pose space using the pre-trained regression head described in 3DmotionBERT. However, temporal misalignment is observed between the generated sequence and the original gait input. This is expected, as GD treats the first-frame feature only as a condition and does not enforce its consistency with the output, assuming potential anomalies in the initial frame. To mitigate this, we apply dynamic time warping (DTW) (Senin 2008) based on Euclidean distance between features, and remove unmatched head or tail segments to obtain temporally aligned raw and reconstructed pairs, denoted as cz_h and cn_h for each h-th hypothesis.

Dual Pathology Disentanglement. We disentangle pathological features in each hypothesis via residual computation. The index h is omitted in this section for brevity. Inspired by prior works such as FSGait (Duan, Wan, and Zhao 2024) and Dual-Conditioned Motion Diffusion (Wang et al. 2025), which suggest that motion anomalies can manifest as both static pose and dynamic motion deviations, we design a DPD module to capture these two pathological patterns.

For pose-level pathology, we directly compute the residual between cn and cz. To suppress potential generation-induced jitters introduced by GD, we adopt a sliding window approach, slicing the sequence into 9-frame clips with a stride of 9 frames. Within each clip, temporal average pooling is performed to obtain a stable representation. This results in a pose-level pathological sequence $sp \in \mathbb{R}^{S\times J\times C}$, where S denotes the number of window clips.

For motion-level pathology, we first derive the dynamic representation d from cz via first-order temporal differencing $d_t = cz^{t+1} - cz^t$. Then, the same residual, slicing, and average pooling are applied to obtain the motion pathological sequence $dm \in \mathbb{R}^{S\times J\times C}$.

Since the number of windows S varies across sequences after DTW temporal alignment, we aggregate pathological information for each h-th hypothesis based on clinical insights that physicians often focus on the most severe pose anomaly and the average dynamic deviation:

$$csp_h = \arg\max_s \text{MPJFM}(sp_s), \tag{10}$$

$$cdm_h = \frac{1}{S}\sum_{s=1}^{S} dm_s \tag{11}$$

where MPJFM computes the mean per-joint feature magnitude of sp_s, indicating the severity of pose pathology.

Method		Venue	Frame	Accuracy ↑	Precision ↑	Recall ↑	F1-Score ↑
Heuristic	Benchmarking	FG 2024	-	<u>0.66</u>	<u>0.68</u>	<u>0.66</u>	<u>0.66</u>
Data-driven	POTR	ICCV 2021	80	0.45	0.49	0.45	0.46
	MixSTE	CVPR 2022	81	0.40	0.41	0.40	0.41
	ST-GCN	JBHI 2022	80	0.49	0.50	0.49	0.48
	MotionBERT	ICCV 2023	81	0.42	0.45	0.42	0.43
	MotionAGFormer	WACV 2024	81	0.42	0.42	0.42	0.42
	PoseFormer-V2*	CVPR 2023	81	0.64	0.65	0.54	0.62
	AGIR*	Arxiv 2025	54	0.51	0.51	0.46	0.58
	3DmotionBERT*	ours	81	0.56	0.56	0.56	0.54
	DPPD*	ours	81	**0.73** 10.6%↑	**0.76** 11.8%↑	**0.73** 10.6%↑	**0.73** 10.6%↑

Table 1: **Quantitative Comparison of Gait Scoring on PDGait.** The best results are highlighted in **bold**, and the second-best results are <u>underlined</u>. Models marked with "*" indicate that their feature extraction networks were finetuned on the PDGait dataset. Frame denotes the sequence length of input used by each model.

Method	G.S.		D.S.	
	Acc.↑	F1-Score↑	Acc.↑	F1-Score↑
OF-DDNet	0.54	0.49	0.69	0.65
KShapeNet	0.54	0.45	0.65	0.55
ST-GCN	0.49	0.44	0.61	0.57
GaitBase	0.43	0.30	0.53	0.42
GaitCLIP	<u>0.68</u>	<u>0.63</u>	<u>0.90</u>	<u>0.84</u>
DPPD	**0.70**	**0.68**	**0.91**	**0.85**

Table 2: **Quantitative Comparison on 3DGait.** "G.S." and "D.S." denote the gait scoring and dementia subtyping. "Acc." is the abbreviation of the accuracy metric.

Quantitative Gait Analysis. The extracted dual-pathological features are further used for gait scoring and dementia subtyping, both formulated as classification problems. We first introduce two learnable weights, $W, M \in \mathbb{R}^{H \times J}$ to adaptively fuse pose and motion pathologies across joints and hypotheses, followed by an average pooling over the joint and hypothesis dimension. Then, a two-layer MLP is adopted as the classification head:

$$y = \text{MLP}(\text{Avg}(W \odot csp + M \odot cdm)) \quad (12)$$

where $\odot$ denotes element-wise multiplication across the joint dimension, and csp is the concatenation of csp_h along the hypothesis axis, as well as cdm. Focal loss (Lin et al. 2017) is employed to address class imbalance in both tasks.

Qualitative Anomaly Highlight. Thanks to the temporal consistency loss in Eq. 2, the magnitude of the pathological features reflects the severity of gait, which is also demonstrated in our experiment. This allows us to localize potentially abnormal joints by identifying those with the highest magnitudes in pathological features. We first apply empirically determined thresholds γ_p and γ_m to filter normal joints in sp_s and dm_s. Subsequently, a Top-K strategy is adopted. K joints with the highest magnitude in sp_s and dm_s are visualized as highlighted colors within the s-th window clip. This enhances our DPPD by (1) providing intuitive visual cues consistent with the quantitative output, and (2) directing clinicians' attention to the severe joints, potentially improving diagnostic confidence and treatment planning.

Experiment

Datasets and Evaluation Metrics

3D Human Pose Datasets. We utilize three widely adopted 3D human pose datasets in this study. Human3.6M (Ionescu et al. 2013) contains 3.6 million video frames performed by 11 actors, captured in a controlled indoor environment. AMASS (Mahmood et al. 2019) aggregates multiple existing motion capture datasets into a unified representation, comprising 344 subjects and over 40 hours of motion videos. CMU Mocap (Lab 2003) provides more than 1,000 action classes across six categories.

Gait Datasets. Two NDDs gait datasets are used in our experiments. PDGait (Shida et al. 2023) is a clinical dataset containing 885 walking sequences collected from 26 individuals diagnosed with PD, using motion capture sensors. 3DGait (Zeng et al. 2023) includes 90 walking sequences from 43 patients diagnosed with PD, AD, or DLB. 3D human pose is extracted from the video by a dedicated 3D pose estimation network (Wang et al. 2023).

Evaluation Metrics. For the PDGait dataset, we perform a 3-class gait scoring (normal-0, slight-1, and mild-2) and evaluate the results using standard classification metrics: Accuracy, Precision, Recall, and F1-Score. For the 3DGait dataset, we evaluate both 4-class gait scoring (normal–0, slight–1, mild–2, and moderate–3) and 3-type dementia subtyping (normal, AD, DLB) tasks. Accuracy and F1-score are reported to quantify the performance. Given the multi-class nature of the tasks, precision, recall, and F1-score are computed as weighted averages based on the class distribution.

Implementation Details

Hyper Parameters. (1) The input 3D pose sequences are uniformly sampled at 30 fps and segmented into clips of $F = 81$ frames with a stride of 9 frames within one motion sequence. (2) For the GD module, the diffusion step is set to $T = 1000$. We set $I = 5$ and $H = 1$ based on ablation studies presented in the Supplementary. (3) The magnitude thresholds are empirically fixed as $\gamma_p = 0.75$ and $\gamma_m = 0.06$, also determined through ablation experiments in the Supplementary.

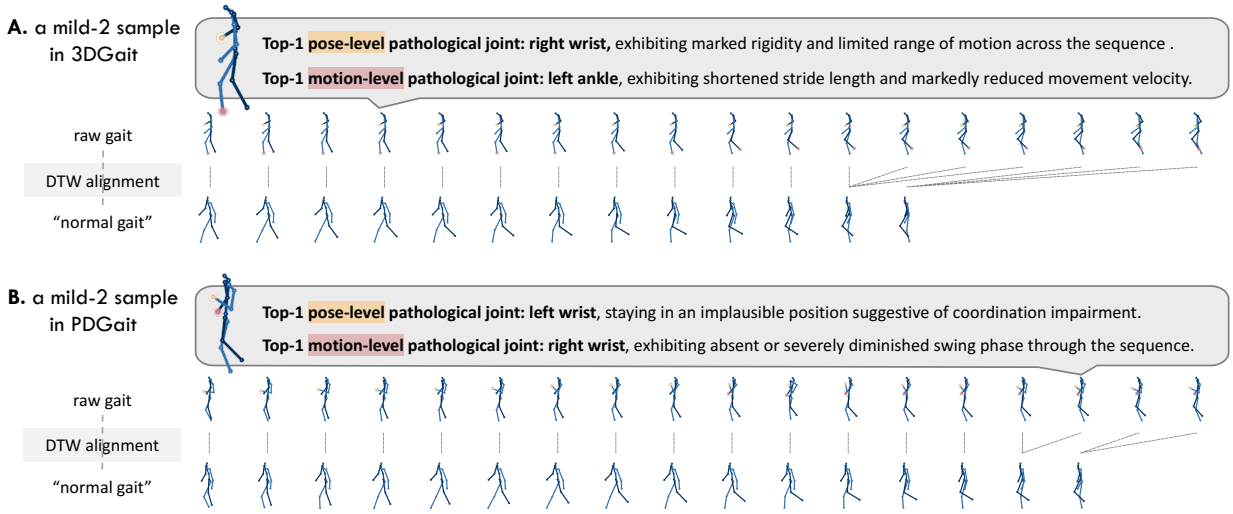

Figure 3: **Visualization of anomaly highlighting.** The first row shows a sample with mild-2 gait severity from the 3DGait dataset, and the second row is from PDGait. Yellow and red indicate the Top-1 anomalous joint with pose-level and motion-level pathology, respectively. An 18-frame segment is shown for clarity. The "normal gait" refers to the 3D pose sequence generated by our GD module from the raw gait representation, with dashed lines denoting their frame-wise temporal alignment.

Training Details. First, 3DmotionBERT is pre-trained for 40 epochs on Human3.6M, AMASS, CMU Mocap, and then finetuned for 20 epochs on PDGait and 3DGait. Second, GD is trained only on normal gait sequences within PDGait and 3DGait for 1000 epochs, with 3DmotionBERT frozen. Finally, we follow the evaluation protocols used in prior works: a Leave-One-Subject-Out Cross-Validation is adopted for PDGait, following Benchmarking (Adeli et al. 2024), and a 10-fold Cross-Validation is employed for 3DGait, following Enhancing (Wang et al. 2024). For each dataset and task, the MLP classifier is independently trained for 20 epochs in each fold. All experiments are conducted on a single NVIDIA 4090 24GB GPU. More detailed settings are in the Supplementary.

Comparison with SOTAs

We compare the proposed DPPD with eight state-of-the-art (SOTA) methods on the PDGait dataset: Benchmarking (Adeli et al. 2024), ST-GCN (Sabo et al. 2022), and AGIR (Wang, Bobenrieth, and Seo 2025), MixSTE (Zhang et al. 2022), MotionBERT (Zhu et al. 2023), PoseFormer-V2 (Zhao et al. 2023), MotionAGFormer (Mehraban, Adeli, and Taati 2024), and PORT (Martínez-González, Villamizar, and Odobez 2021). For the 3DGait dataset, we evaluate DPPD against five SOTA methods: OF-DDNet (Lu et al. 2020), KShapeNet (Friji et al. 2021), ST-GCN (Sabo et al. 2022), GaitBase (Fan et al. 2023), and GaitCLIP (Wang et al. 2024).

Quantitative Results. As shown in Tab. 1, DPPD outperforms all other data-driven methods in the 3-class gait scoring task across all evaluation metrics. Compared with Benchmarking, DPPD achieves relative improvements of 10.6%, 11.8%, 10.6%, and 10.6% in Accuracy, Precision, Recall, and F1-Score, respectively, demonstrating its superior capacity for pathological gait representation in the diagnosis of NDDs. As shown in Tab. 2, DPPD also exhibits competitive performance on the 3DGait dataset. In the dementia subtyping task, DPPD slightly outperforms GaitCLIP, even though it only uses video-based input, in contrast to the multimodal inputs leveraged by GaitCLIP. Compared with other purely video-based methods, DPPD achieves at least 0.22 and 0.20 higher in Accuracy and F1-Score. These results collectively highlight the strong generalization ability of DPPD across different impaired gait analysis tasks in NDDs.

Qualitative Results. Based on the disentangled pathological representation, we highlight Top-K anomalous joints within the window clips to provide interpretability for quantitative results. The value of K can be adjusted as needed. Here, we present the Top-1 results of two mild-2 patients sampled. As illustrated in Fig. 3, the "normal gait" exhibits coordinated gait patterns, including regular arm swings and natural stride lengths, which demonstrates GD's ability to generate normal gait. The first patient displays reduced stride and upper limb stiffness. These abnormalities are accurately captured by our model, with the left ankle identified as Top-1 motion-level anomalous joint, and the right wrist at pose-level. The second patient shows pronounced upper-limb rigidity, characterized by persistent flexion of the left arm and absence of arm swing. Our method correctly identifies the left wrist as the most pose-level pathological joint and the right wrist at motion-level. These results align well with visual observations, validating the reliability of our anomaly highlighting. More visualizations are in the Supplementary material and videos.

Ablation Studies

We conduct ablation studies on PDGait dataset to assess the effect of different components in our DPPD framework.

3D	Finetune	GD	DPD	Accuracy	F1-Score
				0.42	0.43
✓				0.52	0.53
✓	✓			0.56	0.54
✓	✓	✓		0.68	0.66
✓	✓	✓	✓	0.73	0.73

Table 3: **Effect of Each Module.** The first row represents the baseline MotionBERT. "3D" indicates that the backbone is adapted for 3D input. "Finetune" denotes that the 3D MotionBERT is pre-trained on the PDGait dataset.

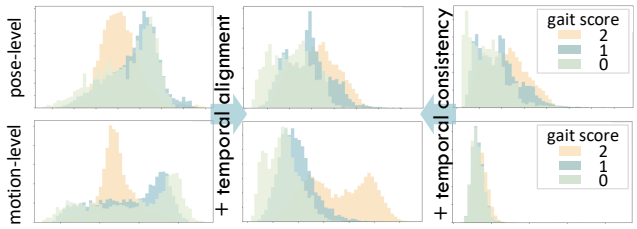

Figure 4: Visualization of the magnitude distribution of pathological representations across different gait severity levels. Green, blue, and yellow correspond to normal-0, slight-1, and mild-2 gait severity, respectively.

Effect of each Module. We evaluate the contribution of each module to the 3-class gait scoring task, as summarized in Tab. 3. Starting from MotionBERT as the baseline, we introduce 3DmotionBERT and pre-train it on 3D human pose datasets. The extracted features are used for gait scoring via MLP. This modification yields over 20% relative improvement in performance, indicating that 3D pose representations offer richer information for gait analysis. Further finetuning it on PDGait and 3DGait provides another gain. Next, we introduce GD to generate normal gait representation and derive residual features for classification. This leads to an additional 21.4% and 22.2% improvement in Accuracy and F1-score, respectively, validating the effectiveness of pathology disentanglement. Finally, incorporating the DPD module to extract dual pathological features after temporal alignment boosts performance by at least 0.05, demonstrating that explicitly modeling static and dynamic pathologies benefits downstream classification.

Effectiveness of Personality Modeling. The GD module aims to reconstruct subject-specific normal gait through personality modeling. However, it is non-trivial to verify whether the learned representation truly captures the personality, given the lack of ground-truth. To evaluate it, we select a subset from PDGait, which only contains data from the subject who has normal gait recordings. A 3-class gait scoring is conducted on this subset. We compare four types of pathological features derived from different sources of normal gait representations: (1) "memory bank –", where it is built as the set of normal features from other subjects without the same subject; (2) "memory bank +", where it only includes features from the same subject; (3) "GD w/o $\mathcal{L}_{per}$", where the normal gait is generated by GD without the personality contrastive loss; and (4) "GD with $\mathcal{L}_{per}$". As shown in Tab. 4, three conclusions are drawn: (1) Pathological features consistently outperform raw features. (2) Personalized normal gait representations are critical, particularly for distinguishing normal from abnormal. In "memory bank -", without personalization, a large number of normal samples are misclassified as 1, leading to notably low Precision. (3) Our GD effectively captures subject-specific characteristics, as its performance lies between the lower bound "memory bank –" and upper bound "memory bank +". Adding $\mathcal{L}_{per}$ further improves personality modeling.

Input	Normal Gait	Accuracy	Precision
raw feature	-	0.53	0.28
pathology feature	memory bank -	0.57	0.32
	memory bank +	0.81	0.65
	GD (w/o $\mathcal{L}_{per}$)	0.64	0.66
	GD (with $\mathcal{L}_{per}$)	0.68	0.66

Table 4: **Effectiveness of Personality Modeling in GD.** "w/o" denotes "without". "-" and "+" indicate whether the memory bank are built from subjects without or within the same subject, respectively.

Effectiveness of Pathological Representation. We propose two essential requirements for the pathological representation: (1) it should exhibit clear separability across different gait severity levels; and (2) its magnitude should be consistent with gait pathology severity, thus supporting qualitative anomaly localization. To meet these criteria, we introduce a temporal consistency loss $\mathcal{L}_{tem}$ and apply a DTW temporal alignment strategy. To assess it, we visualize the distribution of pathological feature magnitudes across different gait severity levels, illustrated in Fig. 4. Comparing the first and second columns, without DTW, the magnitude distribution does not follow the expected progression trend, i.e., greater severity does not consistently lead to greater pathological magnitude. Further, comparing the second and third columns, removing $\mathcal{L}_{tem}$ noticeably reduces the separability of motion pathological features across severity levels. More validations can be seen in the Supplementary and Videos.

Conclusion

In this paper, we propose a Diffusion-based Personalized Pathology Disentanglement model, named DPPD, for jointly quantitative and qualitative impaired gait analysis in NDDs. The framework comprises three key components: 3DMotionBERT for encoding 3D pose sequences, GD for generating personalized normal gait representations, and DPD for disentangling pose-level and motion-level pathological features. Experimental results show that DPPD outperforms existing SOTA methods in both gait scoring and dementia subtyping. Furthermore, the proposed anomaly highlighting visualization provides reliable and interpretable evidence aligned with observation. However, DPPD is currently unable to provide fine-grained descriptions of joints' abnormalities, which we will explore in future work.

Acknowledgements

This work has been funded in part by the NSFC grants 62176156 and the Medical Engineering Cross Research Fund of Shanghai Jiao Tong University (YG2023ZD12).

References

Adeli, V.; Mehraban, S.; Ballester, I.; Zarghami, Y.; Sabo, A.; Iaboni, A.; and Taati, B. 2024. Benchmarking Skeleton-based Motion Encoder Models for Clinical Applications: Estimating Parkinson's Disease Severity in Walking Sequences. In *2024 IEEE 18th International Conference on Automatic Face and Gesture Recognition (FG)*, 1–10. IEEE.

Albuquerque, P.; Verlekar, T. T.; Correia, P. L.; and Soares, L. D. 2021. A spatiotemporal deep learning approach for automatic pathological gait classification. *Sensors*, 21(18): 6202.

Chee, R.; Murphy, A.; Danoudis, M.; Georgiou-Karistianis, N.; and Iansek, R. 2009. Gait freezing in Parkinson's disease and the stride length sequence effect interaction. *Brain*, 132(8): 2151–2160.

Conroy, S. 2008. Emergency room geriatric assessment—urgent, important or both? *Age and ageing*, 37(6): 612–613.

Deng, D.; Ostrem, J. L.; Nguyen, V.; Cummins, D. D.; Sun, J.; Pathak, A.; Little, S.; and Abbasi-Asl, R. 2024. Interpretable video-based tracking and quantification of parkinsonism clinical motor states. *npj Parkinson's Disease*, 10(1): 122.

Duan, B.; Wan, X.; and Zhao, X. 2024. FSGait: Fine Grained Self-Supervised Gait Abnormality Detection. In *Proceedings of the Asian Conference on Computer Vision*, 2248–2264.

Fan, C.; Liang, J.; Shen, C.; Hou, S.; Huang, Y.; and Yu, S. 2023. Opengait: Revisiting gait recognition towards better practicality. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 9707–9716.

Friji, R.; Drira, H.; Chaieb, F.; Kchok, H.; and Kurtek, S. 2021. Geometric deep neural network using rigid and non-rigid transformations for human action recognition. In *Proceedings of the IEEE/CVF international conference on computer vision*, 12611–12620.

Goetz, C. G.; Tilley, B. C.; Shaftman, S. R.; Stebbins, G. T.; Fahn, S.; Martinez-Martin, P.; Poewe, W.; Sampaio, C.; Stern, M. B.; Dodel, R.; et al. 2008. 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.

Guo, C.; Mu, Y.; Javed, M. G.; Wang, S.; and Cheng, L. 2024. Momask: Generative masked modeling of 3d human motions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1900–1910.

He, H.; Zhang, J.; Chen, H.; Chen, X.; Li, Z.; Chen, X.; Wang, Y.; Wang, C.; and Xie, L. 2024. A diffusion-based framework for multi-class anomaly detection. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, 8472–8480.

Ho, J.; Jain, A.; and Abbeel, P. 2020. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33: 6840–6851.

Hsu, W.-C.; Sugiarto, T.; Lin, Y.-J.; Yang, F.-C.; Lin, Z.-Y.; Sun, C.-T.; Hsu, C.-L.; and Chou, K.-N. 2018. Multiple-wearable-sensor-based gait classification and analysis in patients with neurological disorders. *Sensors*, 18(10): 3397.

Hu, T.; Zhang, J.; Yi, R.; Du, Y.; Chen, X.; Liu, L.; Wang, Y.; and Wang, C. 2024. Anomalydiffusion: Few-shot anomaly image generation with diffusion model. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, 8526–8534.

Ionescu, C.; Papava, D.; Olaru, V.; and Sminchisescu, C. 2013. Human3. 6m: Large scale datasets and predictive methods for 3d human sensing in natural environments. *IEEE transactions on pattern analysis and machine intelligence*, 36(7): 1325–1339.

Jiang, B.; Chen, X.; Liu, W.; Yu, J.; Yu, G.; and Chen, T. 2023. Motiongpt: Human motion as a foreign language. *Advances in Neural Information Processing Systems*, 36: 20067–20079.

Jiang, Q.; Liu, J.; Huang, S.; Wang, X.-Y.; Chen, X.; Liu, G.-H.; Ye, K.; Song, W.; Masters, C. L.; Wang, J.; and Wang, Y.-J. 2025. Antiageing strategy for neurodegenerative diseases: from mechanisms to clinical advances. 10(1): 76.

Lab, C. G. 2003. CMU Graphics Lab Motion Capture Database. Technical report, Carnegie Mellon University.

Li, H.; Mehul, A.; Le Kernec, J.; Gurbuz, S. Z.; and Fioranelli, F. 2020. Sequential human gait classification with distributed radar sensor fusion. *IEEE Sensors Journal*, 21(6): 7590–7603.

Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; and Dollár, P. 2017. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, 2980–2988.

Liu, J.; Ma, Z.; Wang, Z.; Zou, C.; Ren, J.; Wang, Z.; Song, L.; Hu, B.; Liu, Y.; and Leung, V. 2025. A survey on diffusion models for anomaly detection. *arXiv preprint arXiv:2501.11430*.

Liu, J.; Wu, K.; Nie, Q.; Chen, Y.; Gao, B.-B.; Liu, Y.; Wang, J.; Wang, C.; and Zheng, F. 2024. Unsupervised continual anomaly detection with contrastively-learned prompt. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, 3639–3647.

Lu, M.; Poston, K.; Pfefferbaum, A.; Sullivan, E. V.; Fei-Fei, L.; Pohl, K. M.; Niebles, J. C.; and Adeli, E. 2020. Vision-based estimation of MDS-UPDRS gait scores for assessing Parkinson's disease motor severity. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 637–647. Springer.

Maaten, L. v. d.; and Hinton, G. 2008. Visualizing data using t-SNE. *Journal of machine learning research*, 9(Nov): 2579–2605.

Mahmood, N.; Ghorbani, N.; Troje, N. F.; Pons-Moll, G.; and Black, M. J. 2019. AMASS: Archive of motion capture as surface shapes. In *Proceedings of the IEEE/CVF international conference on computer vision*, 5442–5451.

Martínez-González, A.; Villamizar, M.; and Odobez, J.-M. 2021. Pose transformers (potr): Human motion prediction with non-autoregressive transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, 2276–2284.

Mehraban, S.; Adeli, V.; and Taati, B. 2024. Motionagformer: Enhancing 3d human pose estimation with a transformer-gcnformer network. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 6920–6930.

Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; and Ommer, B. 2022. High-Resolution Image Synthesis with Latent Diffusion Models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Sabo, A.; Mehdizadeh, S.; Iaboni, A.; and Taati, B. 2022. Estimating parkinsonism severity in natural gait videos of older adults with dementia. *IEEE journal of biomedical and health informatics*, 26(5): 2288–2298.

Sabo, A.; Mehdizadeh, S.; Ng, K.-D.; Iaboni, A.; and Taati, B. 2020. Assessment of Parkinsonian gait in older adults with dementia via human pose tracking in video data. *Journal of neuroengineering and rehabilitation*, 17(1): 97.

Sarapata, G.; Dushin, Y.; Morinan, G.; Ong, J.; Budhdeo, S.; Kainz, B.; and O'Keeffe, J. 2023. Video-based activity recognition for automated motor assessment of Parkinson's disease. *IEEE journal of biomedical and health informatics*, 27(10): 5032–5041.

Senin, P. 2008. Dynamic time warping algorithm review. *Information and Computer Science Department University of Hawaii at Manoa Honolulu, USA*, 855(1-23): 40.

Shida, T. K. F.; Costa, T. M.; de Oliveira, C. E. N.; de Castro Treza, R.; Hondo, S. M.; Los Angeles, E.; Bernardo, C.; dos Santos de Oliveira, L.; de Jesus Carvalho, M.; and Coelho, D. B. 2023. A public data set of walking full-body kinematics and kinetics in individuals with Parkinson's disease. *Frontiers in Neuroscience*, 17: 992585.

Song, J.; Meng, C.; and Ermon, S. 2020. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*.

Tevet, G.; Gordon, B.; Hertz, A.; Bermano, A. H.; and Cohen-Or, D. 2022. Motionclip: Exposing human motion generation to clip space. In *European Conference on Computer Vision*, 358–374. Springer.

Trosch, R. M.; Friedman, J. H.; Lannon, M. C.; Pahwa, R.; Smith, D.; Seeberger, L. C.; O'Brien, C. F.; Lewitt, P. A.; and Koller, W. C. 1998. Clozapine use in Parkinson's disease: a retrospective analysis of a large multicentered clinical experience. *Movement Disorders*, 13(3): 377–382.

Verghese, J.; Lipton, R. B.; Hall, C. B.; Kuslansky, G.; Katz, M. J.; and Buschke, H. 2002. Abnormality of Gait as a Predictor of Non-Alzheimer's Dementia. *New England Journal of Medicine*, 347(22): 1761–1768.

Wang, D.; Bobenrieth, C.; and Seo, H. 2025. AGIR: Assessing 3D Gait Impairment with Reasoning based on LLMs. *arXiv preprint arXiv:2503.18141*.

Wang, D.; Yuan, K.; Muller, C.; Blanc, F.; Padoy, N.; and Seo, H. 2024. Enhancing gait video analysis in neurodegenerative diseases by knowledge augmentation in vision language model. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 251–261. Springer.

Wang, D.; Zouaoui, C.; Jang, J.; Drira, H.; and Seo, H. 2023. Video-based gait analysis for assessing alzheimer's disease and dementia with lewy bodies. In *International Workshop on Applications of Medical AI*, 72–82. Springer.

Wang, H.; Xu, A.; Ding, P.; and Gui, J. 2025. Dual Conditioned Motion Diffusion for Pose-Based Video Anomaly Detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 7700–7708.

Wolfson, L.; Whipple, R.; Amerman, P.; and Tobin, J. N. 1990. Gait assessment in the elderly: a gait abnormality rating scale and its relation to falls. *Journal of gerontology*, 45(1): M12–M19.

Yeh, C.-H.; Wang, J.; Graham, A. D.; Liu, A. J.; Tan, B.; Chen, Y.; Ma, Y.; and Lin, M. C. 2024. Insight: A multimodal diagnostic pipeline using llms for ocular surface disease diagnosis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 711–721. Springer.

Zeng, Q.; Liu, P.; Yu, N.; Wu, J.; Huo, W.; and Han, J. 2023. Video-based quantification of gait impairments in Parkinson's disease using skeleton-silhouette fusion convolution network. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31: 2912–2922.

Zhang, H.; Wang, Z.; Zeng, D.; Wu, Z.; and Jiang, Y.-G. 2025a. DiffusionAD: Norm-Guided One-Step Denoising Diffusion for Anomaly Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(8): 7140–7152.

Zhang, J.; Tu, Z.; Yang, J.; Chen, Y.; and Yuan, J. 2022. Mixste: Seq2seq mixed spatio-temporal encoder for 3d human pose estimation in video. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 13232–13242.

Zhang, T.; Gao, L.; Li, X.; and Gao, Y. 2025b. DZAD: Diffusion-based Zero-shot Anomaly Detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 10131–10138.

Zhao, Q.; Zheng, C.; Liu, M.; Wang, P.; and Chen, C. 2023. Poseformerv2: Exploring frequency domain for efficient and robust 3d human pose estimation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 8877–8886.

Zhu, W.; Ma, X.; Liu, Z.; Liu, L.; Wu, W.; and Wang, Y. 2023. Motionbert: A unified perspective on learning human motion representations. In *Proceedings of the IEEE/CVF international conference on computer vision*, 15085–15099.