Cross-Spatiotemporal Graph Convolution Networks for Skeleton-Based Parkinsonian Gait MDS-UPDRS Score Estimation

Haoyu Tian, Haiyun Li, Wenjing Jiang, Xin Ma, *Senior Member, IEEE*, Xiang Li, Hanbo Wu, and Yibin Li, *Member, IEEE*

***Abstract*— Gait impairment in Parkinson's Disease (PD) is quantitatively assessed using the Movement Disorder Society Unified Parkinson's Disease Rating Scale (MDS-UPDRS), a well-established clinical tool. Objective and efficient PD gait assessment is crucial for developing interventions to slow or halt its advancement. Skeleton-based PD gait MDS-UPDRS score estimation has attracted increasing interest in improving diagnostic efficiency and objectivity. However, previous works ignore the important cross-spacetime dependencies between joints in PD gait. Moreover, existing PD gait skeleton datasets are very small, which is a big issue in deep learning-based gait studies. In this work, we collect a sizable PD gait skeleton dataset by multi-view Azure Kinect sensors. The collected dataset contains 102 PD patients and 30 healthy older adults. In addition, gait data from 16 young adults (aged 24-50 years) are collected to further examine the effect of age on PD gait assessment. For skeleton-based automatic PD gait analysis, we propose a novel cross-spatiotemporal graph convolution network (CST-GCN) to learn complex features of gait patterns. Specifically, a gait graph labeling strategy is designed to assemble and group cross-spacetime neighbors of the root node according to the spatiotemporal semantics of the gait skeleton. Based on this strategy, the CST-GCN module explicitly models the cross-spacetime dependencies among joints. Finally, a dual-path model is presented to realize the modeling and fusion of spatial, temporal, and cross-spacetime gait features. Extensive experiments validate the effectiveness of our method on the collected dataset.**

***Index Terms*— Parkinsonian gait, quantitative assessment, graph convolutional network, skeleton-based data.**

I. Introduction

PARKINSON'S Disease (PD) is the second most common neurodegenerative disease, which affects millions of people around the world [1], [2], [3], [4]. The prevalence of PD increases with age, and it is rare in people under 50 years old [5], [6]. In terms of pathophysiology, PD is caused by reduced amount of dopamine in the brain. Dopamine is a vital neurotransmitter in the execution of spontaneous movements [7], [8]. Hence, dopamine deficiency usually provokes symptoms of significant gait disturbance [9], [10]. Since the severity of gait impairment marks the progression of PD disease, accurate assessment of PD gait is crucial for adjusting clinical treatments [11]. Currently, clinicians assess the severity of PD gait impairments by part III of the Movement Disorder Society Unified Parkinson's Disease Rating Scale (MDS-UPDRS) [12]. Due to the requirement for clinical visits, the PD gait MDS-UPDRS score based on manual inspection is performed infrequently in daily life, making it prone to overlooking gait changes in patients. Moreover, this method is semi-subjective and inefficient. Therefore, there is a pressing need to develop an automated technology capable of objectively estimating the MDS-UPDRS score of PD gait.

In recent years, various gait analysis systems have been investigated for automatic quantification of PD gait impairments using multi-camera motion capture (Mocap) systems [13], foot force sensors [16], [17], and multiple inertial measurement units (IMU) systems [14], [15]. While Mocap and force sensors provide accurate lower limbs' kinematics and dynamics for gait assessment, they require complicated sensor protocols and calibration. It is difficult to collect gait data using these complex and expensive sensors. Additionally, the invasive nature of wearable IMU can affect the naturalness of a patient's gait. Since depth sensors, such as Kinect, offer a non-intrusive and less expensive platform for capturing human motion, it has gained popularity in gait

Manuscript received 5 July 2023; revised 10 September 2023 and 21 November 2023; accepted 6 January 2024. Date of publication 10 January 2024; date of current version 19 January 2024. This work was supported in part by the Basic Research Key Development Program of Shandong Province under Grant ZR2019ZD07, in part by the National Natural Science Foundation of China-Regional Innovation Development Joint Fund Project under Grant U21A20486, in part by the Fundamental Research Funds for the Central Universities under Grant 2022JC011, and in part by the Science and Technology Innovation 2030-"New Generation Artificial Intelligence" Major Project under Grant 2020AAA0108903. *(Haoyu Tian and Haiyun Li contributed equally to this work.) (Corresponding authors: Wenjing Jiang; Xin Ma.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Research Ethics Board at Qilu Hospital of Shandong University, under Application No. KYLL-202203-028.

Haoyu Tian, Xin Ma, Xiang Li, Hanbo Wu, and Yibin Li are with the Center of Robotics School of Control Science and Engineering, Shandong University, Jinan 250061, China (e-mail: tianhaoyu@mail.sdu.edu.cn; maxin@sdu.edu.cn; lixiang0814@mail.sdu.edu.cn; wuhanbo@sdu.edu.cn; liyb@sdu.edu.cn).

Haiyun Li and Wenjing Jiang are with the Qilu Hospital of Shandong University, Jinan 250012, China (e-mail: lihaiyun1122@qiluhospital.com; jiangwenjing@qiluhospital.com).

Digital Object Identifier 10.1109/TNSRE.2024.3352004

TABLE I
MDS-UPDRS STANDARD FOR PROGRESSIVE PD GAIT IMPAIRMENTS

Score		Symptom Manifestation
0	Normal	No problems.
1	Slight	Independent walking with minor gait impairments.
2	Mild	Independent walking but with substantial gait impairments.
3	Moderate	Requires assistance device for safe walking but not a person.
4	Severe	Cannot walk at all or only with another person's assistance.

analysis [18], [19]. Many studies validate that the Kinect sensor exhibits acceptable accuracy for examining gait [20], [21], [22], [23]. Moreover, the Kinect sensor enables real-time 3D human skeleton tracking which alleviates the effect of human appearance differences and lighting variations. Deep learning techniques can accurately recognize human movements represented by skeleton data [24], [25], [26]. Supported by recent advances in depth sensors and computer vision, it is possible to automatically estimate PD gait MDS-UPDRS score in daily life.

We refer to the clinical scoring scheme of the MDS-UPDRS to describe PD gait impairment in detail. The scores range from 0, indicating no gait impairments, to 4 for patients unable to walk independently, as depicted in Table I. Scores 0 to 2 indicate a progressive decrease in mobility, characterized by reduced stride amplitude and foot lift. These types of PD gait are not easy to be distinguished from each other. In contrast, PD gaits scored more than 2 are easy to identify since these patients cannot walk independently. The clinical doctors assess the patient's gait by timed up-and-go (TUG) tests, which consist of six sub-tasks, i.e., Sit, Sit-to-Stand, Walk, Turn, Walk-Back, and Sit-Back. Since patients with scores of 3 and 4 are associated with a high risk of falls, we only collected TUG test data for Parkinsonian gait with scores of 0, 1, and 2 in this work.

Many studies have proposed skeleton-based method to achieve an automatic PD gait MDS-UPDRS score estimation. Lu et al. instructed 55 Parkinson's patients to perform the gait exams and proposed an ordinal-focal double-feature double-motion network (OF-DDNet) to quantify PD gait impairments [28]. The OF-DDNet concatenates coordinate vectors of all joints to form a single feature vector per frame, which cannot effectively learn the latent spatial features. In works by Sabo et al. [27] and Guo et al. [47], spatial-temporal graph convolution network (ST-GCN) was employed to effectively model spatial and temporal gait features. Sabo et al. [27] collected 399 natural gait videos from 53 elderly individuals, where only 14 participants exhibited drug-induced Parkinsonism gait. In [47], 441 gait examination video clips of 142 patients were collected for the assessment of PD gait. These advances have demonstrated the feasibility of skeleton-based PD gait quantitative analysis. However, the above PD gait datasets generally suffer from limitations such as few subjects, insufficient gait data, and a single viewpoint. The absence of large-scale datasets limits the development of deep models for quantitative assessment of PD gait. The gait circle is segmented into four phases: swing phase, initial double-support phase, single-support phase, and end double-support phase [30]. The cross-spacetime dependencies between joints are critical for characterizing PD gait severity. However, ST-GCN-based methods have limitations in capturing complex cross-spatiotemporal dependencies due to obstructed information flow. There is no research on the direct modeling of cross-spacetime gait features for PD gait MDS-UPDRS score estimation.

To solve the above issues, we used Azure Kinect to collect a sizeable PD gait skeleton dataset which contains 148 participants (102 Parkinson's patients and 46 healthy people). Two experts conducted MDS-UPDRS scoring for 102 Parkinson's patients. The gait of 46 healthy individuals was assigned a default score of 0. The healthy individuals consist of 16 younger participants below the age of 50 and 30 older individuals who match the age and gender of the PD group. Three Azure Kinect cameras placed at different positions were employed to capture TUG depth videos of participants. Subsequently, 3D skeleton data from six different viewpoints was extracted from the recorded videos using the corresponding SDK. The multi-view dataset is more consistent with the requirements of realizing automatic gait assessment in the home environment. To our knowledge, our collected dataset is the largest in the field of 3D skeleton-based PD gait MDS-UPDRS score estimation. Such a sizable and diverse dataset is conducive to the development of deep learning-based PD gait analysis.

Moreover, we propose a novel cross-spatiotemporal graph convolution network (CST-GCN) to enhance the modeling of complex PD gait features. First, we design a gait graph and corresponding cross-spacetime neighbor graph labeling function according to the semantics of the gait skeleton. The labeling strategy is scalable in spatial and temporal dimensions of the gait graph, which divides the cross-spacetime neighbor nodes of the root node into a fixed number of subsets. Based on the cross-spacetime labeling strategy, CST-GCN is introduced to learn the cross-spacetime dependencies of joints. The CST-GCN adopts unshared weights in spatial and temporal dimensions, which enhances the capability of modeling the cross-spacetime gait features. Finally, a dual-path model is constructed to model and fuse the spatial, temporal, and cross-spacetime features for PD gait MDS-UPDRS score estimation. Extensive experiments on the collected dataset demonstrate the effectiveness of the proposed CST-GCN and dual-path model for PD gait assessment.

In summary, the contributions of our work are summarized as follows:

1) We collect a multi-viewpoint PD gait skeleton dataset from 148 participants, which contains 102 Parkinson's patients and 46 healthy controls. To our knowledge, the collected dataset is the largest in the field of skeleton-based automatic PD gait assessment. Such a sizable multi-view dataset is conducive to the development of deep learning-based PD gait MDS-UPDRS score estimation.
2) We propose a novel CST-GCN to learn the cross-spacetime features for PD gait MDS-UPDRS

score estimation. A cross-spacetime graph labeling strategy is introduced to partition the cross-spacetime neighbor gait graph according to gait skeleton semantics. Based on the labeling strategy, CST-GCN adopts unshared weights to enhance the capability of modeling cross-spacetime gait features. Moreover, a dual-path model is constructed to simultaneously learn spatial, temporal, and cross-spacetime gait features.

3) Extensive experiments on the collected dataset validate the effectiveness of the proposed method.

II. Related Works

A. Gait Dataset

In the field of abnormal gait recognition, many works have collected and publicly released vision-based abnormal gait datasets. Paiement et al. [31] gathered gait data encompassing normal, Parkinson's, and Stroke conditions, with the abnormal gaits of the latter two conditions simulated by five healthy adults. Chaaraoui et al. [32] utilized Kinect v2 to collect gait data from 7 volunteers, who simulated four abnormal gaits: right knee injury, left knee injury, right foot dragging, and left foot-dragging. In another walking gait dataset [33], 9 healthy volunteers were guided to simulate the pathological asymmetric gait by padding sole (5 cm, 10 cm, and 15 cm soles) and attaching weight (4kg weights). In the multi-modal gait symmetry (MMGS) dataset [22], 27 healthy volunteers were asked to simulate abnormal gait by using insoles or keeping their knees stiff. Because the collection of patient data requires formal permissions and involves privacy protection, the above abnormal gait datasets were obtained from healthy volunteers.

For skeleton-based PD gait MDS-UPDRS score estimation, Lu et al. collected videos of the TUG test from 55 participants, and first quantified movement-linked impairments for assessing PD severity on MDS-UPDRS [28], [34]. Cao et al. recorded the gait video of 18 Parkinson's patients and 42 healthy for detecting freezing of gait [37]. Sabo et al. [27] collected 399 natural gait videos from 53 elderly individuals, where only 14 participants exhibited drug-induced Parkinsonism gait. In a recent study, a two-view PD gait dataset was collected for PD gait analysis. However, it comprises only 9 patients with Parkinson's disease [35]. The above PD gait datasets generally suffer from limitations such as few subjects, insufficient gait data, and a single viewpoint. The absence of a sizeable dataset limits the exploration of deep learning-based PD gait assessment and hinders application potential. In this work, we collect a sizeable PD gait 3D skeleton dataset to fully explore the problems existing in automatic PD gait MDS-UPDRS score estimation, such as the gait feature distribution and age effects.

B. Skeleton-Based Gait Analysis

Vision-based gait analysis relies primarily on RGB video [36], [37], [38], [39] and skeleton sequences [22], [28], [35], [49], [50]. In comparison to videos, skeleton data exhibits robustness to environmental changes. Many works adopted hand-crafted gait features based on prior knowledge to achieve skeleton-based abnormal gait recognition [31], [40], [41]. However, the hand-crafted gait features have limited representation and generalization because of missing important clues. Khokhlova et al. [22] proposed a bidirectional Long Short-Term Memory (LSTM) network for evaluating gait symmetry. Guo et al. [42] conducted a comparison between the support vector machine (SVM) and LSTM network, finding that the LSTM network improved the accuracy of gait classification. Lu et al. [28] presented an ordinal focal convolution network to estimate the MDS-UPDRS scores of PD gait. The work in [35] introduced a comprehensive quantitative comparison of 16 diverse traditional machine and deep learning algorithms, which demonstrates the potential of the deep learning-based framework for neurological gait dysfunction prediction. The above methods concatenate coordinate vectors of all joints to form a single feature vector per frame. Since the gait skeleton is represented essentially as a series of non-Euclidean graphs, these methods cannot effectively learn the latent spatial dependencies in skeleton joints.

For effectively modeling spatial gait features, Liu et al. proposed a symmetry-driven hyper feature GCN for the skeleton-based gait identity recognition [50]. In [49], a GCN-based approach, integrating higher-order inputs and residual networks, was introduced for gait identity recognition. Many works adopted ST-GCN [24] for skeleton-based abnormal gait analysis [27], [43], [44]. The work in [44] designed a spatiotemporal attention module to enhance the capability of ST-GCN for learning the fine-grained gait features. However, the above methods have limitations in achieving unobstructed cross-spacetime information flow for capturing complex spatial-temporal dependencies of joints.

In skeleton-based action recognition, MS-G3D introduced dense cross-spacetime edges as skip connections for direct information propagation across the spatial-temporal graph [25]. However, MS-G3D adopts shared weight in the time dimension, which limits the modeling of cross-spacetime features. In contrast to MS-G3D, our CST-GCN module has two important distinctions: (1) The labeling function groups the cross-spacetime neighbor node according to the spatiotemporal semantics of the gait skeleton. (2) The CST-GCN deploys unshared weights in spatial and temporal dimensions, which enhances the capacity for learning cross-spacetime dependencies among joints.

III. Data Collection

A. Assumptions and Limitations of the System

Skeleton data of the patient in the TUG test can be used for automatic PD gait MDS-UPDRS score estimation [27], [47]. In the data collection system, three Azure Kinect cameras placed at different locations are used to capture the 3D skeleton gait data of participants. The Kinect records depth video at a resolution of 320 by 288 pixels, operating at a frequency of 30 Hz. The operating range of the Azure Kinect camera is 0.5-5.46 meters. The 3D skeleton data are delivered from the depth video by the corresponding SDK. Participants were instructed to perform the TUG test on a 3-meter length carpet. The setup for data collection is shown in Fig. 1. In the confined spatial setting, the deployment of multi-view

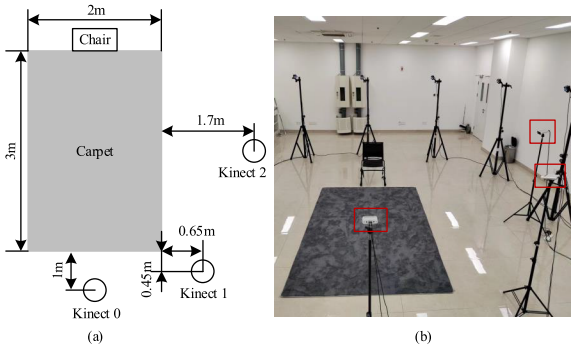

Fig. 1. (a) The schematic displays a bird eye view of the equipment arrangement. Three Azure Kinect sensors, Kinect 0, Kinect 1, and Kinect 2, are placed at different locations. (b) Photographs show the layout of Azure Kinect sensors highlighted by the red rectangle.

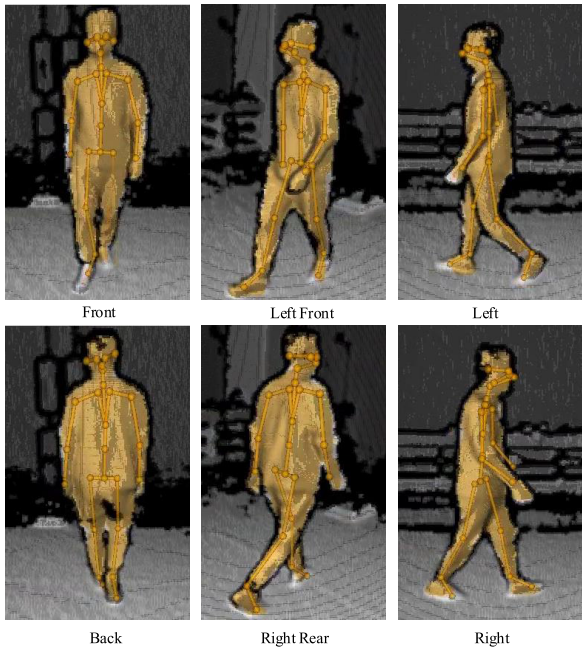

Fig. 2. Six views of the walk, i.e. Front (F), Back (B), Left Front (LF), Right Rear (RR), Left (L), and Right (R).

data acquisition more aptly conforms to the requirements of conducting Parkinson's disease (PD) gait analysis within a home environment.

In practical application, the gait acquisition area is relatively small due to the operational range limitations of the Azure Kinect camera. Patients are required to walk within the designated area, ensuring the absence of obstacles or other individuals.

B. Data Collection

The dataset was collected by the Neurology Department of Qilu Hospital from March 2022 to April 2023. All procedures were approved by the Research Ethics Board at Qilu Hospital of Shandong University, which prohibits public dataset release due to patient privacy constraints. We collected skeleton data of the TUG test from 148 participants, which contains 102 PD patients and 46 healthy controls. Walks from PD patients were scored on the MDS-UPDRS scales by two specialists, and

TABLE II
SCORE DISTRIBUTION AND RELATED INFORMATION OF THE 148 SUBJECTS

	Socre	0	1	2
PD	N	-	54	48
	Gender(F/M)	-	29/25	19/29
	Age(mean±std)	-	64.7 ± 8.0	66.2 ± 8.0
OC	N	30	-	-
	Gender(F/M)	16/14	-	-
	Age(mean±std)	62.7 ± 7.4	-	-
YC	N	16	-	-
	Gender(F/M)	6/10	-	-
	Age(mean±std)	35.0 ± 9.1	-	-

the control group had a default score of 0. Because of the high severity of gait impairment, patients with scores 3 and 4 are at high risk of falling. We only collected gait data from PD patients with scores of 1 and 2. All patients were not taking anti-Parkinsonian medication or "off" anti-Parkinsonian medication state. Moreover, we collected the gait data of 16 healthy adults (YC) younger than 50 years old in the control group to explore the effect of age on PD gait MDS-UPDRS score estimation. The remaining 30 controls were healthy older adults (OC) whose age and gender were matched with PD groups. Score distribution and related information of the 148 gait participants are provided in Tab. II. Each PD subject was recorded twice, while each control subject was recorded four times. The TUG test comprises two gait-related subtasks: walking and walking back. By employing three Azure Kinect sensors, gait data from six views (Front, Back, Left Front, Right Rear, Left, and Right) can be collected in a single TUG test, as depicted in Fig. 2. In total, the dataset contains 2314 sequences of gait skeletons. Compared with the existing studies on skeleton-based PD gait MDS-UPDRS score estimation, our dataset encompasses the largest number of PD patients and multi-view gait data. This augments the exploration of deep learning methods for automatic PD gait analysis.

IV. METHODS

In this section, we initially provide a preliminary overview of spatial GCN and temporal GCN. Subsequently, we formulate the gait graph and its corresponding graph labeling function. Finally, a CST-GCN is proposed to learn cross-spacetime gait features. Additionally, we propose a dual-path model to learn and fuse the spatial, temporal, and cross-spacetime gait features for PD gait MDS-UPDRS score estimation.

A. Preliminaries

The skeleton sequence captured by Kinect is a temporal series of human joint positions represented as 3D coordinates in each frame. It can be represented as a spatiotemporal non-Euclidean graph, as illustrated in Fig. 3. ST-GCN [24] serves as a foundational framework for extracting motion features from non-Euclidean skeleton graphs, encompassing sequential spatial GCN and temporal GCN steps.

1) Spatial GCN: In contrast to the 2D image convolution, it is trickier to define the weight function $\mathbf{W}$ for spatial GCN. The spatial GCN utilizes a graph labeling process to map all neighboring vertices into a fixed number of K subsets, each associated with a unique learnable weight vector. The layer-wise propagation rule of spatial GCN is given as follows:

$$\mathbf{F}_{out} = \sigma\left(\sum_{k}^{K} \hat{\mathbf{A}}_k \mathbf{F}_{in} \mathbf{W}_k\right), \tag{1}$$

where $\mathbf{F}_{in} \in \mathbb{R}^{N \times C_{in}}$ is the input feature map, $\mathbf{F}_{out} \in \mathbb{R}^{N \times C_{out}}$ is the output feature map, $\mathbf{W}_k \in \mathbb{R}^{C_{in} \times C_{out}}$ is a learnable weight vector, C_{in} and C_{out} are numbers of channels, $\hat{\mathbf{A}}_k \in \mathbb{R}^{N \times N}$ is normalized adjacency matrix of $\mathbf{A}_k$, N is the number of joints, and $\sigma(\cdot)$ denotes an activation function.

2) Temporal GCN: The basic temporal GCN is a classical temporal-only convolution layer (TCN). Concretely, TCN performs a $\Gamma \times 1$ convolution on the output feature map $\mathbf{F}_{out}$ in the spatial GCN (1), Γ denotes the kernel size for the temporal dimension. The TCN only involves the same joint on the inter-frame.

B. Gait Data and Processing

1) Gait Skeleton Data: We used Azure Kinect DK to obtain 3D coordinates (X, Y, Z) of the joints. In skeleton-based gait analysis, many methods only use the coordinate of lower limb joints as input [29], [35], [44]. This strategy is implemented to prevent the introduction of redundant information from additional nodes. In this work, the gait skeleton data is composed of the coordinates of nine joints in the lower limbs, as shown in Fig. 3(a). In our methodology, we employ the gait skeleton sequence $S \in \mathbb{R}^{T \times N \times C_0}$ as input, where T is the sequence length, $N = 9$ is the number of joints in gait skeleton, $C_0 = 3$ is the channel number of 3D coordinates.

2) Data Pre-Processing: Each gait skeleton is normalized by subtracting the positions of the 0-base of the spine joint,

$$\hat{S} = \left[S_{(:,i,:)} - S_{(:,0,:)} | i \in [0, 1, \ldots, N-1]\right], \tag{2}$$

where $\hat{S} \in \mathbb{R}^{T \times N \times C_0}$ is the normalized gait skeleton data, $S_{(:,i,:)} \in \mathbb{R}^{T \times 1 \times C_0}$ is the positions of the $i-th$ joint, and $S_{(:,0,:)} \in \mathbb{R}^{T \times 1 \times C_0}$ is the positions of the 0-base of the spine joint.

3) Gait Graph: In the spatial dimension, the joints within one frame are connected according to the human body structure, as illustrated in Fig. 3(a). When employing a factorized spatial and temporal GCN to model gait graph sequences, each joint is only connected to the same joint in the consecutive frame, as shown in Fig. 3(b). In the case of utilizing a cross-spacetime GCN for modeling gait graph sequences, each joint is connected to the cross-spacetime neighbor joints in consecutive frames, as shown in Fig. 3(c).

C. Spatial Graph Labeling Function

On the spatial gait graph, we define the spatial neighbor set $B_s(v_{ti}) = \left\{v_{tj} \,|\, d(v_{tj}, v_{ti}) \le D\right\}$ of a node v_{ti}. Here $d(v_{tj}, v_{ti})$ gives the shortest length of any path from v_{tj} to v_{ti}. D is less than or equal to the diameter of the spatial gait graph. The set B_s can be interpreted as the spatial receptive field of the spatial GCN. For the gait skeleton graph depicted in Fig. 1 (b), we define the temporal neighbor set $B_t(v_{ti}) = \{v_{\tau i} \,||\tau - t| \le \Gamma/2\}$ of a node v_{ti}.

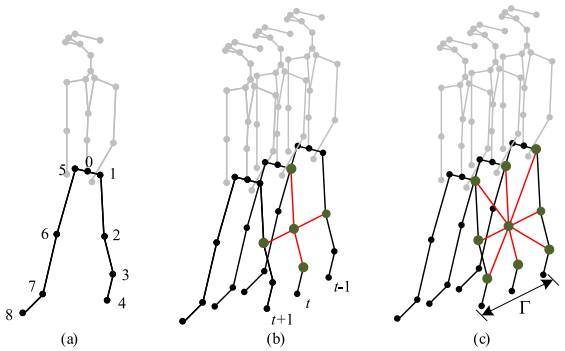

Fig. 3. (a) Spatial gait graph. The IDs of the joints are 0-base of the spine, 1-left hip, 2-left knee, 3-left ankle, 4-left foot, 5-right hip, 6-right knee, 7-right ankle, and 8-right foot. (b) Each joint is connected to the same joint in the consecutive frame. (c) Each joint is connected to the cross-spacetime neighbor joints in consecutive frames.

If the aggregation of long-range ($D > 1$) spatial information is performed using higher-order polynomials $\hat{\mathbf{A}}^D$ of the adjacency matrix, it may introduce bias towards local regions and nodes with higher degrees. MS-G3D [25] introduced the disentangling neighborhoods labeling function $l^s(v_{tj}, v_{ti}) = d(v_{tj}, v_{ti})$ to address the above problem. We follow this idea to construct a spatial labeling function for the proposed gait graph. To encode the geometric characteristics of the gait graph, joints with the same distance from the root node v_i are further divided into the left group and the right group. The left group denotes the neighboring nodes that are closer to the 4-left foot node than the root node, i.e. $d(v_j, v_4) < d(v_i, v_4)$. Conversely, the right group denotes $d(v_j, v_4) \ge d(v_i, v_4)$, where $d(v_j, v_4) = d(v_i, v_4)$ refers to the root node itself. This strategy is inspired by the symmetry of gait. The spatial labeling function can be formulated as:

$$l^s(v_{tj}, v_{ti}) = \begin{cases} 2d(v_{tj}, v_{ti}) - 1 & \text{if } d(v_j, v_4) < d(v_i, v_4), \\ 2d(v_{tj}, v_{ti}) & \text{else,} \end{cases} \tag{3}$$

where $v_{tj} \in B_s(v_{ti})$, and $0 \le l^s(v_{tj}, v_{ti}) \le 2D$. The proposed labeling function divides the spatial neighbor set $B_s(v_{ti})$ into $2D+1$ subsets. The corresponding adjacency matrix is defined for each subset as follows:

$$\mathbf{A}^s_{(k)}[i, j] = \begin{cases} 1 & \text{if } l^s(v_{tj}, v_{ti}) = k, \\ 0 & \text{else,} \end{cases} \tag{4}$$

where $\mathbf{A}^s_{(k)} \in \mathbb{R}^{N \times N}$, and $0 \le k \le 2D$. Based on the proposed spatial graph labeling function (3), we arrive at the factorized spatial and temporal GCN operator as follows:

$$\mathbf{X}^{(l+1)} = TCN\left(\sigma\left(\sum_{k=0}^{2D} \hat{\mathbf{A}}^s_{(k)} \mathbf{X}^{(l)} \mathbf{W}^{(l)}_{(k)}\right)\right), \tag{5}$$

where $\mathbf{X}^{(l)} \in \mathbb{R}^{T \times N \times C^{(l)}}$ is the input gait feature, $\mathbf{X}^{(l+1)} \in \mathbb{R}^{T \times N \times C^{(l+1)}}$ is the gait feature generated by factorized spatial

and temporal GCN, $C^{(l)}$ and $C^{(l+1)}$ are the number of feature channels, $\mathbf{W}_{(k)}^{(l)} \in \mathbb{R}^{C^{(l)} \times C^{(l+1)}}$ is the learnable parameter, $\hat{\mathbf{A}}_{(k)}^{s} = {D_{(k)}^{s}}^{-1} \mathbf{A}_{(k)}^{s}$ is the normalized adjacency matrix of $\mathbf{A}_{(k)}^{s}$, $D_{(k)}^{s}$ is the degree matrix of $\mathbf{A}_{(k)}^{s}$, $\sigma(\cdot)$ is an activation function, and TCN is employed as temporal GCN operator.

D. Cross-Spacetime Graph Convolution Network (CST-GCN)

The factorized spatial and temporal GCN hinders the cross-spacetime information flow, which can not effectively learn the complex PD gait features. We propose the CST-GCN to model complex cross-spacetime gait features for the MDS-UPDRS score estimation of PD gait. First, a temporal window of size Γ is slid over the gait graph sequence with zero padding to obtain T cross-spacetime subgraphs. Each subgraph encompasses all nodes across Γ frames, as shown in Fig. 3 (c). By extending the spatial receptive field B_s to consecutive frames on the cross-spacetime subgraph, the cross-spacetime neighbor set of the root node v_{ti} is defined as $B_{st}(v_{ti}) = \left\{ v_{\tau j} \left| |\tau - t| < \Gamma/2 \& d\left(v_{\tau j}, v_{\tau i}\right) \leq D \right. \right\}$. B_{st} can be regarded as the cross-spacetime receptive field of the CST-GCN. To construct the weight function of the CST-GCN, we further extend the spatial graph labeling function $l^{s}\left(v_{tj}, v_{ti}\right)$ to the cross-spacetime graph. This results in the development of the cross-spacetime graph labeling function, designed as follows:

$$l^{st}\left(v_{\tau j}, v_{ti}\right) = l^{s}\left(v_{j}, v_{i}\right) + \left(\begin{matrix} \tau - t \\ + \lfloor \Gamma/2 \rfloor \end{matrix} \right) (2D+1), \quad (6)$$

where $\lfloor \cdot \rfloor$ is the round-down function, $v_{\tau j} \in B_{st}(v_{ti})$, and $0 \leq l^{st}\left(v_{\tau j}, v_{ti}\right) < \Gamma(2D+1)$. The cross-spacetime graph labeling function divides the cross-spacetime neighbor set $B_{st}(v_{ti})$ into $\Gamma(2D+1)$ subsets. On the cross-spacetime subgraph, adjacency matrixes are defined as follows:

$$\mathbf{A}_{(k)}^{st}\left[i, \left(\begin{matrix} \tau - t + \\ \lfloor \Gamma/2 \rfloor \end{matrix} \right) N + j \right] = \begin{cases} 1 & \text{if } l^{st}(v_{\tau j}, v_{ti}) = k, \\ 0 & \text{else,} \end{cases} \quad (7)$$

where $\mathbf{A}_{(k)}^{st} \in \mathbb{R}^{N \times \Gamma N}$, and $0 \leq k < \Gamma(2D+1)$. Based on the proposed cross-spacetime graph labeling function (6), we formulate the CST-GCN operator as follows:

$$\mathbf{X}^{(l+1)} = \sigma\left(\sum_{k=0}^{\Gamma(2D+1)} \hat{\mathbf{A}}_{(k)}^{st} sw\left(\mathbf{X}^{(l)}\right) \mathbf{W}_{(k)}^{(l)} \right), \quad (8)$$

where $\mathbf{X}^{(l)} \in \mathbb{R}^{T \times N \times C^{(l)}}$ is the input gait feature, $\mathbf{X}^{(l+1)} \in \mathbb{R}^{T \times N \times C^{(l+1)}}$ is the gait feature generated by CST-GCN, $C^{(l)}, C^{(l+1)}$ are the number of feature channels, $\mathbf{W}_{(k)}^{(l)} \in \mathbb{R}^{C^{(l)} \times C^{(l+1)}}$ is the learnable parameter, $\hat{\mathbf{A}}_{(k)}^{st} = {\mathbf{D}_{(k)}^{st}}^{-1} \mathbf{A}_{(k)}^{st}$ is the normalized adjacency matrix of $\mathbf{A}_{(k)}^{st}$, $\mathbf{D}_{(k)}^{st}$ is the degree matrix of $\mathbf{A}_{(k)}^{st}$, $\sigma(\cdot)$ is an activation function, and $sw : \mathbf{X}^{(l)} \in \mathbb{R}^{T \times N \times C^{(l)}} \rightarrow \mathbf{X}_{\Gamma}^{(l)} \in \mathbb{R}^{T \times \Gamma N \times C^{(l)}}$ is the sliding window operator. In contrast to the MS-G3D [25], the CST-GCN employs unshared weights $\mathbf{W}_{(k)}^{(l)}$ in temporal dimensions to enhance the modeling of cross-spacetime gait features.

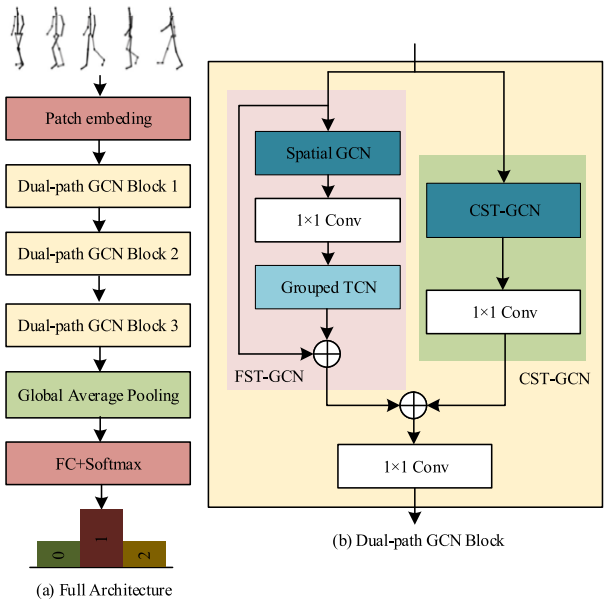

Fig. 4. Model architecture. (a) The overall architecture. The model is composed of 3 layers of dual-path GCN blocks, the output channels of the layers are 64, 128, 256. The global average pooling, fully connected layer, and SoftMax are used to obtain the classification score. (b) The dual-path GCN block. The dual-path GCN block deploys the FST-GCN pathway and CST-GCN pathway to simultaneously capture complex spatial, temporal, and cross-spacetime dependencies of joints. The FST-GCN pathway contains spaital-GCN, point-wise convolution, and grouped TCN. The CST-GCN pathway includes a CST-GCN layer followed by a point-wise convolutional layer. The outputs from two pathways are aggregated by a 1×1 convolutional layer to enhance the fusion of multi-path features.

E. Model Architecture

In this section, we design a dual-path model to learn and fuse spatial, temporal, and cross-spacetime gait features for automatic PD gait MDS-UPDRS score estimation. The overall model architecture is illustrated in Fig. 4 (a). On a high level, it contains one patch embedding block and three dual-path GCN blocks, followed by a global average pooling layer and a SoftMax classifier. First, patch embeddings of size $T \times N \times C^{(1)}$ are computed by a TCN layer in the patch-embedding block. After that, the patch embeddings are passed through three dual-path GCN blocks. The output from dual-path GCN block 1 manifests as a feature map with dimensions $T \times N \times C^{(1)}$. Subsequently, dual-path GCN block 2 yields a feature map sized $T \times N \times 2C^{(1)}$, while the output of dual-path GCN block 3 constitutes a feature map of dimensions $T \times N \times 4C^{(1)}$. $C^{(1)}$ is set to 64.

As shown in Fig. 4 (b), each dual-path GCN block deploys two pathways to simultaneously capture temporal, spatial, and cross-spacetime dependencies of joints, and the outputs from two pathways are aggregated by a 1×1 Conv layer. The FST-GCN pathway includes spatial-GCN layer, 1×1 Conv layer, and channel-grouped TCN layer. The 1×1 Conv layer is inserted in the middle of the spatial-GCN layer and the TCN layer. The kernel size of grouped TCN is 1×7, and the temporal stride is set to 1. The CST-GCN pathway includes a CST-GCN layer and a 1×1 Conv layer. To reduce the parameters of the model, we use grouped pointwise

convolution to replace the pointwise linear projection in spatial GCN and CST-GCN. For all grouped convolutions, the number of channels in each group is 16. To simplify the expression, all the default BN or ReLU after convolution are omitted in Fig. 4.

V. Experiments

A. Experiment Setting

All experiments are conducted on the PyTorch deep learning framework and run on an Nvidia RTX3090 GPU with 24GB memory. All models are trained with Stochastic gradient descent (SGD) with Nesterov momentum (0.9). Cross-entropy is selected as the loss function to back-propagate gradients. 48 frames of skeleton data randomly selected from each gait sequence are down-sampled to 16 frames as input. Our models are trained with batch size 8, an initial learning rate of 0.001, and a weight decay of 0.0005. The learning rate is divided by 10 at the 20th epoch. The training process ended at the 25th epoch. The parameter setting of each experiment is the same, and a random number seed is set to ensure complete reproducibility of the experimental results.

B. Evaluation Metrics

Accuracy (Acc), precision (Prec), recall (Rec), f1-score (F1), and area under the ROC curve (AUC) are used as evaluation metrics to evaluate the classification results. These metrics are defined as follows:

$$\begin{aligned} \text{Acc} &= \frac{\text{TP}+\text{TN}}{\text{TP}+\text{FN}+\text{TN}+\text{FP}}, \\ \text{Prec} &= \frac{\text{TP}}{\text{TP}+\text{FP}}, \\ \text{Rec} &= \frac{\text{TP}}{\text{TP}+\text{FN}}, \\ \text{F1} &= \frac{2\times\text{Prec}\times\text{Rec}}{\text{Prec}+\text{Rec}}. \end{aligned} \tag{9}$$

where TP, TN, FP, and FN represent the number of true positive, true negative, false positive, and false negative samples, respectively.

C. Validation Strategies

The experiments are performed by a participant-based five-fold cross-validation. The dataset is randomly divided into five fixed independent folds; four folds are used for training, and the remaining one is used for independent testing. Samples of each participant are restricted to be in exactly one fold so that the training set and the testing set do not have samples from the same participant. We set two evaluation benchmarks, i.e. Cross-Subject (CS) and Cross-Subject-View (CSV). In the CS evaluation, we split the 148 subjects into training and testing groups according to the participant-based five-fold cross-validation. For CSV evaluation, the training dataset was collected through Kinects 0 and 1, while the testing dataset was obtained using Kinect 2, following the participant-based five-fold cross-validation strategy.

TABLE III
Comparison With Baseline and Ablated Methods

Methods	Acc	Prec	Rec	F1	AUC
ST-GCN [24] (baseline)	0.663	0.640	0.626	0.629	0.789
FST-GCN (25joint)	0.664	0.627	0.622	0.615	**0.814**
FST-GCN (our)	**0.670**	**0.656**	**0.633**	**0.637**	0.804
G3D [25] (baseline)	0.661	0.654	0.627	0.633	0.802
CST-GCN (our)	**0.672**	**0.656**	**0.633**	**0.636**	**0.803**
Dual-path model (our)	**0.677**	**0.666**	**0.641**	**0.647**	0.804

D. Ablation Experiments

In this section, we conduct ablation experiments on the CS-benchmark to validate the effectiveness of our proposed method.

1) Gait Graph and Spatial Graph Labeling Function: The individual FST-GCN pathway is adopted to validate the effectiveness of the proposed gait graph and corresponding labeling function. We take the ST-GCN [24] as the baseline in this section. We also test the FST-GCN pathway based on the whole body skeleton graph and the spatial configuration partitioning strategy [24], namely FST-GCN (25 joints). The spatial and temporal receptive fields are set to $D = 1$ and $\Gamma = 7$. As shown in Table III, the FST-GCN (25 joints) outperforms the ST-GCN by 0.025 AUC, which validates the effectiveness of the proposed FST-GCN pathway. The proposed FST-GCN with the proposed gait graph and labeling function outperforms the FST-GCN (25 joints) by 0.006 accuracy and 0.022 F1-score. These results further validate the effectiveness of the proposed gait graph and corresponding spatial graph labeling function.

2) CST-GCN: In this section, we validate the effectiveness of the individual CST-GCN pathway. The G3D pathway of Ms-GCN [25] is used as a baseline. These experiments are conducted on our proposed gait graph, the spatial and temporal receptive fields are set to $D = 2$ and $\Gamma = 3$. The CST-GCN outperforms the G3D by 0.011 accuracy and 0.003 F1-score, as shown in Table III. These results validate that the proposed CST-GCN enhances the capacity for learning cross-spacetime gait features.

3) Dual-Path Model: Finally, we validate the effectiveness of the dual-path model, as shown in Tab III. The proposed dual-path model achieves accuracy of 0.677 and F1-score of 0.647, which exceeds the CST-GCN pathway by 0.005 accuracy and 0.011 F1-score. These results fully illustrate the effectiveness of the dual-path model.

E. Comparison With Baseline Methods

Machine learning models (K-Nearest Neighbors (KNN), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM) with poly kernels, and Extreme Gradient Boosting (XGBoost)) and deep learning models (OF-DDNet [28], AGS-GCN [44], GaitGraph2 [49], ST-GCN [24], and Ms-G3D [25]) are selected as baselines to evaluate the performance of our dual-path model. As shown in Table IV, although our method exhibits a slightly lower AUC than MS-G3D on CS protocol, it achieves the best accuracy

TABLE IV
COMPARISON WITH BASELINE METHODS ON CS BENCHMARK. BEST RESULTS ARE IN BOLD AND SECOND BEST ARE UNDERLINED

Method	Acc	Prec	Rec	F1	AUC
KNN	0.526	0.501	0.515	0.499	0.687
DT	0.461	0.449	0.447	0.448	0.585
RF	0.550	0.503	0.503	0.489	0.680
SVM	0.555	0.536	0.538	0.537	0.714
XGBoost	0.611	0.577	0.577	0.574	0.760
OF-DDNet [28], [34]	0.626	0.597	0.586	0.584	0.737
ST-GCN [24]	0.663	0.640	0.626	0.629	0.789
AGS-GCN [44]	0.665	0.637	0.634	0.631	0.797
MS-G3D [25]	0.668	0.637	0.637	0.636	**0.810**
GaitGraph2 [49]	0.666	0.628	0.631	0.621	0.804
Our model	**0.677**	**0.666**	**0.641**	**0.647**	0.804

TABLE V
COMPARISON WITH BASELINE METHODS ON CSV BENCHMARK. BEST RESULTS ARE IN BOLD AND SECOND BEST ARE UNDERLINED

Method	Acc	Prec	Rec	F1	AUC
KNN	0.532	0.503	0.518	0.502	0.685
DT	0.474	0.445	0.452	0.447	0.589
RF	0.537	0.481	0.488	0.459	0.654
SVM	0.547	0.514	0.523	0.516	0.701
XGBoost	0.589	0.551	0.547	0.536	0.745
OF-DDNet [28], [34]	0.584	0.571	0.561	0.565	0.725
ST-GCN [24]	0.690	0.684	0.653	0.660	0.798
AGS-GCN [44]	0.655	0.631	0.620	0.614	0.780
MS-G3D [25]	0.676	0.649	0.640	0.639	0.808
GaitGraph2 [49]	0.684	0.667	0.652	0.656	0.794
Our model	**0.700**	**0.679**	**0.664**	**0.658**	**0.818**

TABLE VI
MODEL COMPLEXITY OF THE PROPOSED MODEL AND OTHER GCN-BASED METHODS

Method	Params (M)	FLOPs (G)	Training time (s/epoch)	Inference computational time(s)
ST-GCN [24]	3.07	1.09	4.88	0.0048
AGS-GCN [44]	1.39	**0.05**	3.86	0.0033
MS-G3D [25]	2.79	1.63	36.29	0.0260
GaitGraph2 [49]	2.99	0.99	5.45	0.0043
Our model	**0.33**	0.12	**3.52**	**0.0031**

and F1-score, which outperforms MS-G3D by 0.009 and 0.011, respectively. Moreover, as shown in Table V, the proposed method surpasses other models across all evaluation metrics on CSV protocol, which validates its robustness to viewpoint changes.

F. Model Complexity

We provide a comparison of the complexity (parameters (Params) and floating point operations (FLOPs)) and running speed (training phase and inference computational time) between our proposed model and other GCN-based models, as shown in Table VI. The Params and FLOPs in our model are notably lower compared to ST-GCN, MS-G3D, and GaitGraph2. While AGS-GCN [44] has fewer FLOPs than our model, its accuracy is significantly worse than our method, as shown in Tables IV and V. Moreover, our model accelerates the training process and reduces inference computational time. Overall, we provide an energy-efficient method for Parkinson's gait assessment.

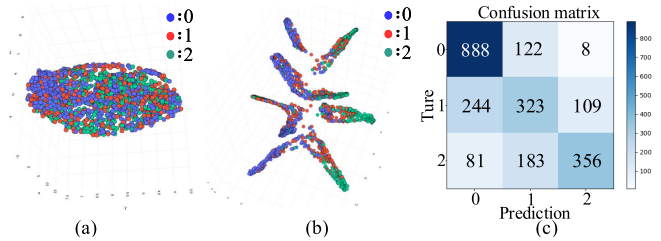

Fig. 5. Visualization of gait features and confusion matrix on the CS benchmark. (a) UMAP embedding clusters of the raw test data. (b) UMAP embedding clusters of the representative features generated by our model. (c) Confusion matrices on CS benchmark.

G. Feature Visualization and Confusion Matrices

To analyze the results of the proposed model in more detail, we visualize the distribution of the test data and the representative features (256 dimensional) generated by our model on the CS benchmark. Uniform Manifold Approximation and Projection (UMAP) [45] is adopted to perform the dimension reduction of the gait skeleton data and the representative features. It can be seen from Fig. 5(a) that the feature embeddings of raw gait skeleton data are highly scattered in nature. There are no pre-defined cluster patterns for the different gait impairments, which makes the task of predicting the UPDRS scores difficult. The learned representative features are more clustered compared to the raw gait data, as shown in Fig. 5(b). These results explain the effectiveness of our model. However, the distribution of learned representative features gradually transitions from 0 to 2, and there is no clear boundary between different gait impairments. It can also be seen from the confusion matrix (Fig. 5(c)) that the test samples are easily misclassified into classes adjacent to the ground truth. The reason for this phenomenon is that gait impairments are symptoms of continuous development, which also brings difficulties in achieving automatic scoring of PD gait.

H. Age Effects

Since elderly people without PD could also experience gait disturbance with age, we assess the effect of age for our method on the CS benchmark. Fig. 6 shows the age distribution and predicted score of these participants with different MDS-UPDRS gait scores. As shown in Fig 6 (a), participants with MDS-UPDRS score 0 but misclassified as 2 are all over 50 years old, indicating that our model is biased by gait impairments associated with aging. To further validate the age effects on PD gait recognition, we conduct an analysis of variance (ANOVA) [46] on the age of subjects with different predicted scores. For the subjects with MDS-UPDRS scores 1 and 2, the ANOVA test returns p-value of 5.7×10^{-5} and 6.2×10^{-17}, which indicates that there are significant age differences across different predicted scores. In conclusion, senility-induced gait impairments can interfere with the automatic analysis of PD gait. Disentangling PD gait features and senility-related gait

TABLE VII
COMPARISON WITH EXISTING RELATED PD GAIT ASSESSMENT STUDIES

Author/Year	Subjects/gaits	Viewpoint	Features	Methods	Acc	Prec	Rec	F1
Sabo et al./2020 [48]	14/398	Front	16 3D features	Logistic regression	0.621	0.607	0.614	0.610
Guo et al./2021 [47]	142/441	Front	2D skeleton	ST-GCN-based method	0.657	0.654	0.651	0.649
Sabo et al./2022 [27]	53/399	Front	3D skeleton	ST-GCN-based method	0.529	0.553	0.670	0.520
Zeng et al./2023 [51]	80/480	Right+left	2D skeleton	ST-GCN-based method	-	0.663	0.683	0.670
			Skeleton+silhouette	ST-GCN+VGG	0.713	0.713	0.710	0.710
Our model	**148/2314**	**Six-viewpoint**	3D skeleton	CST-GCN method	0.677	0.666	0.641	0.647

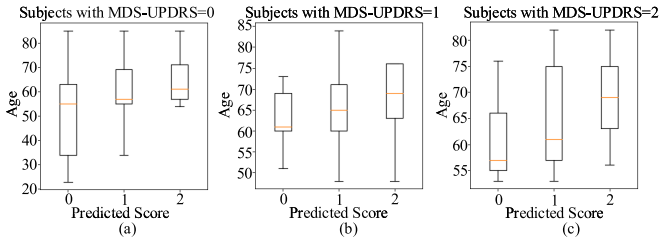

Fig. 6. Subjects with different MDS-UPDRS scores are plotted with their predicted score and their age, respectively. (a) Subjects with ground-truth MDS-UPDRS scores of 0. ANOVA test returns a p-value = 4.7×10^{-9} on the OC group (30 healthy older), confirming a significant age difference between the subjects with predicted scores 0, 1 and 2. (b) Subjects with ground-truth MDS-UPDRS scores of 1. ANOVA test returns a p-value = 5.7×10^{-5}, confirming an insignificant age difference across the groups with different predicted scores. (c) Subjects with ground-truth MDS-UPDRS scores of 2. ANOVA test returns a p-value = 6.2×10^{-17}, which validates that there are significant age differences across the groups with different predicted scores.

features is of great significance to improve the PD gait analysis performance. We leave this to future works.

I. Comparison With Existing Related Studies

Many PD gait assessment studies have demonstrated the challenges in accurately estimating UPDRS gait scores [27], [47], [48], as shown in Table VII. Compared to these existing studies, our dataset includes more participants, gait samples, and viewpoints. The proposed model achieves better accuracy than other skeleton-based methods [27], [47], [48], which validates the effectiveness of the proposed CST-GCN model for PD gait assessment. Moreover, Zeng et al. [51] improved the accuracy of PD gait score estimation by fusing skeleton and silhouette data. However, their gait dataset is considerably smaller compared to our multi-view gait dataset. The above qualitative analysis demonstrates the excellence of our multi-view PD gait dataset and the proposed CST-GCN method.

VI. CONCLUSION

In this paper, we use Azure Kinect sensors to collect a sizable multi-view PD gait skeleton dataset. To the best of our knowledge, this dataset is the largest in the field of 3D skeleton-based PD gait MDS-UPDRS score estimation. The CST-GCN is proposed to effectively learn cross-spacetime gait features that are important for recognizing PD gait impairments. We further construct a dual-path model to simultaneously model and fuse spatial, temporal, and cross-spacetime features of PD gait. We extensively evaluate our model on the collected PD gait dataset. The results demonstrate significantly improved performance in automatic PD gait analysis. However, since PD gait impairments are a series of progressive symptoms, it is very difficult to find the appropriate classification boundary even using the deep model. Moreover, senility-induced gait impairments could affect the automatic PD gait analysis. We argue that how to decouple PD gait features and senility-related gait features is a meaningful research direction in automatic PD gait classification. Another limitation of this study is that the gait data collection was performed in the hospital. Although we use multiple-view Azure Kinect sensors and carpets to mimic the daily environment, the white coat effect may affect the gait performance of patients. It may be very difficult but necessary to collect a large-scale gait dataset from patients in their home environment.

REFERENCES

[1] E. R. Dorsey, T. Sherer, M. S. Okun, and B. R. Bloem, "The emerging evidence of the Parkinson pandemic," *J. Parkinson's Disease*, vol. 8, no. 1, pp. 3–8, Dec. 2018.
[2] A. E. Lang and A. M. Lozano, "Parkinson's disease," *New England J. Med.*, vol. 339, no. 16, pp. 1130–1143, 1998.
[3] T. Ellis, J. T. Cavanaugh, G. M. Earhart, M. P. Ford, K. B. Foreman, and L. E. Dibble, "Which measures of physical function and motor impairment best predict quality of life in Parkinson's disease?" *Parkinsonism Rel. Disorders*, vol. 17, no. 9, pp. 693–697, Nov. 2011.
[4] L. V. Kalia and A. E. Lang, "Parkinson's disease," *Lancet*, vol. 386, pp. 896–912, Aug. 2015.
[5] O. Riedel, D. Bitters, U. Amann, E. Garbe, and I. Langner, "Estimating the prevalence of Parkinson's disease (PD) and proportions of patients with associated dementia and depression among the older adults based on secondary claims data," *Int. J. Geriatric Psychiatry*, vol. 31, no. 8, pp. 938–943, Aug. 2016.
[6] W. Poewe et al., "Parkinson disease," *Nat. Rev. Dis. Prim.*, vol. 3, no. 1, pp. 1–21, 2017.
[7] J.-H. Baik et al., "Parkinsonian-like locomotor impairment in mice lacking dopamine D2 receptors," *Nature*, vol. 377, no. 6548, pp. 424–428, Oct. 1995.
[8] H. Bernheimer, W. Birkmayer, O. Hornykiewicz, K. Jellinger, and F. Seitelberger, "Brain dopamine and the syndromes of Parkinson and Huntington clinical, morphological and neurochemical correlations," *J. Neurological Sci.*, vol. 20, no. 4, pp. 415–455, Dec. 1973.
[9] M. J. Armstrong and M. S. Okun, "Diagnosis and treatment of Parkinson disease: A review," *Jama*, vol. 323, no. 6, pp. 548–560, 2020.
[10] I. Benatru, M. Vaugoyeau, and J.-P. Azulay, "Postural disorders in Parkinson's disease," *Neurophysiologie Clinique/Clinical Neurophysiol.*, vol. 38, no. 6, pp. 459–465, 2008.
[11] Y. Guo, J. Yang, Y. Liu, X. Chen, and G.-Z. Yang, "Detection and assessment of Parkinson's disease based on gait analysis: A survey," *Frontiers Aging Neurosci.*, vol. 14, p. 837, Aug. 2022.
[12] C. G. Goetz 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*, vol. 23, no. 15, pp. 2129–2170, 2008.

[13] H. Park, S. Shin, C. Youm, S.-M. Cheon, M. Lee, and B. Noh, "Classification of Parkinson's disease with freezing of gait based on 360° turning analysis using 36 kinematic features," *J. NeuroEng. Rehabil.*, vol. 18, no. 1, pp. 1–18, Dec. 2021.

[14] B. Shi, A. Tay, W. L. Au, D. M. L. Tan, N. S. Y. Chia, and S.-C. Yen, "Detection of freezing of gait using convolutional neural networks and data from lower limb motion sensors," *IEEE Trans. Biomed. Eng.*, vol. 69, no. 7, pp. 2256–2267, Jul. 2022.

[15] R. Liu et al., "A wearable gait analysis and recognition method for Parkinson's disease based on error state Kalman filter," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 8, pp. 4165–4175, Aug. 2022.

[16] I. El Maachi, G.-A. Bilodeau, and W. Bouachir, "Deep 1D-convnet for accurate Parkinson disease detection and severity prediction from gait," *Exp. Syst. Appl.*, vol. 143, Apr. 2020, Art. no. 113075.

[17] E. Balaji, D. Brindha, V. K. Elumalai, and R. Vikrama, "Automatic and non-invasive Parkinson's disease diagnosis and severity rating using LSTM network," *Appl. Soft Comput.*, vol. 108, Sep. 2021, Art. no. 107463.

[18] F. Chen et al., "Gait acquisition and analysis system for osteoarthritis based on hybrid prediction model," *Computerized Med. Imag. Graph.*, vol. 85, Oct. 2020, Art. no. 101782.

[19] S. Bei, Z. Zhen, Z. Xing, L. Taocheng, and L. Qin, "Movement disorder detection via adaptively fused gait analysis based on Kinect sensors," *IEEE Sensors J.*, vol. 18, no. 17, pp. 7305–7314, Sep. 2018.

[20] T. M. Guess, S. Razu, A. Jahandar, M. Skubic, and Z. Huo, "Comparison of 3D joint angles measured with the Kinect 2.0 skeletal tracker versus a marker-based motion capture system," *J. Appl. Biomechanics*, vol. 33, no. 2, pp. 176–181, Apr. 2017.

[21] M. Eltoukhy, J. Oh, C. Kuenze, and J. Signorile, "Improved Kinect-based spatiotemporal and kinematic treadmill gait assessment," *Gait Posture*, vol. 51, pp. 77–83, Jan. 2017.

[22] M. Khokhlova, C. Migniot, A. Morozov, O. Sushkova, and A. Dipanda, "Normal and pathological gait classification LSTM model," *Artif. Intell. Med.*, vol. 94, pp. 54–66, Mar. 2019.

[23] J. Guo, Q. Zhang, H. Chai, and Y. Li, "Obtaining lower-body Euler angle time series in an accurate way using depth camera relying on optimized Kinect CNN," *Measurement*, vol. 188, Jan. 2022, Art. no. 110461.

[24] S. Yan, Y. Xiong, and D. Lin, "Spatial temporal graph convolutional networks for skeleton-based action recognition," in *Proc. AAAI Conf. Artif. Intell.*, vol. 32, 2018, pp. 1–9.

[25] Z. Liu, H. Zhang, Z. Chen, Z. Wang, and W. Ouyang, "Disentangling and unifying graph convolutions for skeleton-based action recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 140–149.

[26] K. Xu, F. Ye, Q. Zhong, and D. Xie, "Topology-aware convolutional neural network for efficient skeleton-based action recognition," in *Proc. AAAI Conf. Artif. Intell. (AAAI)*, vol. 36, 2022, pp. 2866–2874.

[27] A. Sabo, S. Mehdizadeh, A. Iaboni, and B. Taati, "Estimating parkinsonism severity in natural gait videos of older adults with dementia," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 5, pp. 2288–2298, May 2022.

[28] M. Lu et al., "Quantifying Parkinson's disease motor severity under uncertainty using MDS-UPDRS videos," *Med. Image Anal.*, vol. 73, Oct. 2021, Art. no. 102179.

[29] X. Gu, Y. Guo, F. Deligianni, B. Lo, and G.-Z. Yang, "Cross-subject and cross-modal transfer for generalized abnormal gait pattern recognition," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 2, pp. 546–560, Feb. 2021.

[30] T. Wang et al., "Recognizing parkinsonian gait pattern by exploiting fine-grained movement function features," *ACM Trans. Intell. Syst. Technol.*, vol. 8, no. 1, pp. 1–22, Jan. 2017.

[31] A. Paiement, L. Tao, M. Camplani, S. Hannuna, D. Damen, and M. Mirmehdi, "Online quality assessment of human motion from skeleton data," in *Proc. Brit. Mach. Vis. Conf.*, 2014, pp. 153–166.

[32] A. A. Chaaraoui, J. R. Padilla-López, and F. Flórez-Revuelta, "Abnormal gait detection with RGB-D devices using joint motion history features," in *Proc. 11th IEEE Int. Conf. Workshops Autom. Face Gesture Recognit. (FG)*, vol. 7, May 2015, pp. 1–6.

[33] T.-N. Nguyen and J. Meunier, "Walking gait dataset: Point clouds, skeletons and silhouettes," Image Process. Lab, DIRO, Univ. Montreal, Tech. Rep., 1379, 2018.

[34] M. Lu et al., "Vision-based estimation of MDS-UPDRS gait scores for assessing Parkinson's disease motor severity," in *Medical Image Computing and Computer Assisted Intervention—MICCAI*. Lima, Peru: Springer, 2020, pp. 637–647.

[35] R. Kaur, R. W. Motl, R. Sowers, and M. E. Hernandez, "A vision-based framework for predicting multiple sclerosis and Parkinson's disease gait dysfunctions—A deep learning approach," *IEEE J. Biomed. Health Informat.*, vol. 27, no. 1, pp. 190–201, Jan. 2023.

[36] L. C. Guayacán and F. Martínez, "Visualising and quantifying relevant parkinsonian gait patterns using 3D convolutional network," *J. Biomed. Informat.*, vol. 123, Nov. 2021, Art. no. 103935.

[37] X. Cao et al., "Video based shuffling step detection for parkinsonian patients using 3D convolution," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 641–649, 2021.

[38] R. Wang et al., "Gait recognition via gait period set," *IEEE Trans. Biometrics, Behav., Identity Sci.*, vol. 5, no. 2, pp. 183–195, Apr. 2023.

[39] T. T. Verlekar, H. De Vroey, K. Claeys, H. Hallez, L. D. Soares, and P. L. Correia, "Estimation and validation of temporal gait features using a markerless 2D video system," *Comput. Methods Programs Biomed.*, vol. 175, pp. 45–51, Jul. 2019.

[40] A. Procházka, O. Vysata, M. Valis, O. Tupa, M. Schätz, and V. Marík, "Bayesian classification and analysis of gait disorders using image and depth sensors of Microsoft Kinect," *Digit. Signal Process.*, vol. 47, pp. 169–177, Dec. 2015.

[41] O. Tupa et al., "Motion tracking and gait feature estimation for recognising Parkinson's disease using MS Kinect," *Biomed. Eng. OnLine*, vol. 14, no. 1, pp. 1–20, Dec. 2015.

[42] Y. Guo, F. Deligianni, X. Gu, and G.-Z. Yang, "3-D canonical pose estimation and abnormal gait recognition with a single RGB-D camera," *IEEE Robot. Autom. Lett.*, vol. 4, no. 4, pp. 3617–3624, Oct. 2019.

[43] D. Mehta et al., "Towards automated and marker-less Parkinson disease assessment: Predicting UPDRS scores using sit-stand videos," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2021, pp. 3836–3844.

[44] H. Tian, X. Ma, H. Wu, and Y. Li, "Skeleton-based abnormal gait recognition with spatio-temporal attention enhanced gait-structural graph convolutional networks," *Neurocomputing*, vol. 473, pp. 116–126, Feb. 2022.

[45] L. McInnes, J. Healy, and J. Melville, "UMAP: Uniform manifold approximation and projection for dimension reduction," 2018, *arXiv:1802.03426*.

[46] F. J. Anscombe, "The validity of comparative experiments," *J. Roy. Stat. Soc. Ser. A*, vol. 111, no. 3, pp. 181–211, 1948.

[47] R. Guo, X. Shao, C. Zhang, and X. Qian, "Multi-scale sparse graph convolutional network for the assessment of parkinsonian gait," *IEEE Trans. Multimedia*, vol. 24, pp. 1583–1594, 2022.

[48] A. Sabo, S. Mehdizadeh, K.-D. Ng, A. Iaboni, and B. Taati, "Assessment of parkinsonian gait in older adults with dementia via human pose tracking in video data," *J. NeuroEng. Rehabil.*, vol. 17, no. 1, pp. 1–10, Dec. 2020.

[49] T. Teepe, J. Gilg, F. Herzog, S. Hörmann, and G. Rigoll, "Towards a deeper understanding of skeleton-based gait recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2022, pp. 1568–1576.

[50] X. Liu, Z. You, Y. He, S. Bi, and J. Wang, "Symmetry-driven hyper feature GCN for skeleton-based gait recognition," *Pattern Recognit.*, vol. 125, May 2022, Art. no. 108520.

[51] Q. Zeng, P. Liu, N. Yu, J. Wu, W. Huo, and J. Han, "Video-based quantification of gait impairments in Parkinson's disease using skeleton-silhouette fusion convolution network," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 2912–2922, 2023.