Cross-Covariate Gait Recognition: A Benchmark

Shinan Zou[1], **Chao Fan**[2,3], **Jianbo Xiong**[1], **Chuanfu Shen**[2,4], **Shiqi Yu**[2,3], **Jin Tang**[1*]

[1]School of Automation, Central South University
[2]Department of Computer Science and Engineering, Southern University of Science and Technology
[3]Research Institute of Trustworthy Autonomous System, Southern University of Science and Technology
[4]The University of Hong Kong
{zoushinan, jianbo_x, tjin}@csu.edu.cn, {12131100, 11950016}@mail.sustech.edu.cn, yusq@sustech.edu.cn

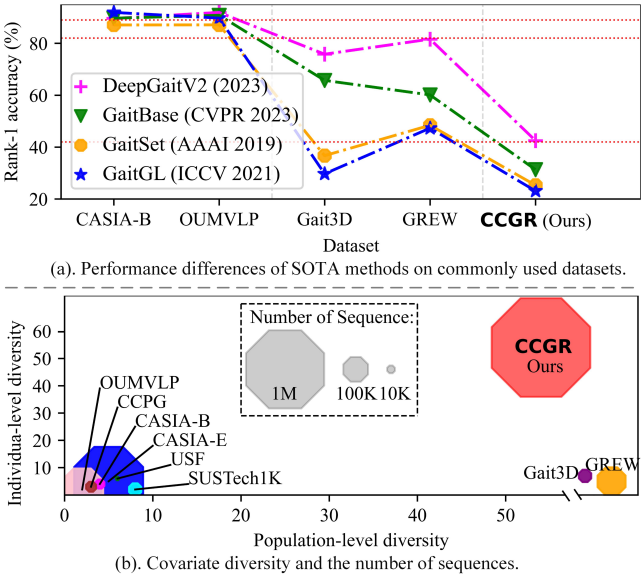
Abstract

Gait datasets are essential for gait research. However, this paper observes that present benchmarks, whether conventional constrained or emerging real-world datasets, fall short regarding covariate diversity. To bridge this gap, we undertake an arduous 20-month effort to collect a cross-covariate gait recognition (CCGR) dataset. The CCGR dataset has 970 subjects and about 1.6 million sequences; almost every subject has 33 views and 53 different covariates. Compared to existing datasets, CCGR has both population and individual-level diversity. In addition, the views and covariates are well labeled, enabling the analysis of the effects of different factors. CCGR provides multiple types of gait data, including RGB, parsing, silhouette, and pose, offering researchers a comprehensive resource for exploration. In order to delve deeper into addressing cross-covariate gait recognition, we propose parsing-based gait recognition (ParsingGait) by utilizing the newly proposed parsing data. We have conducted extensive experiments. Our main results show: 1) Cross-covariate emerges as a pivotal challenge for practical applications of gait recognition. 2) ParsingGait demonstrates remarkable potential for further advancement. 3) Alarmingly, existing SOTA methods achieve less than 43% accuracy on the CCGR, highlighting the urgency of exploring cross-covariate gait recognition. Link: https://github.com/ShinanZou/CCGR.

Introduction

Gait recognition aims to use physiological and behavioral characteristics extracted from walking videos to certify individuals' identities. Compared to other biometric modalities, such as face, fingerprints, and iris, gait patterns have the distinct advantage of being extracted from a distance in uncontrolled environments. These strengths place gait recognition as an effective solution for security applications.

In the latest literature, the research on gait recognition is developing rapidly, with the evaluation benchmark developing from early indoor to outdoor environments. During this remarkable journey, most representative gait models (Chao et al. 2019; Lin, Zhang, and Yu 2021) boasting historical progress have unexpectedly performed unsatisfactory results when faced with emerging challenges posed by real-world gait datasets such as GREW (Zhu et al. 2021) and Gait3D (Zheng et al. 2022). Surprisingly, successive works (Fan et al. 2023b,a) quickly address this performance gap to a large extent, rekindling the promise of gait recognition for practical applications, as illustrated in Figure 1(a). However, this paper argues that the gait recognition task is much more challenging than these datasets have defined.



(a). Performance differences of SOTA methods on commonly used datasets.

(b). Covariate diversity and the number of sequences.

Figure 1: Differences between CCGR and other datasets. Population-level diversity is roughly quantified by the count of covariate categories present within the whole dataset. Correspondingly, individual-level diversity is measured by the count of covariate categories for each subject. Here, the population-level diversity of Gait3D and GREW is rich, but the exact amount is unknown due to the wild scenarios.

In general, previous indoor gait datasets often require subjects repeatedly walk along fixed paths while introducing variations in clothing and carrying. This approach yields controllable and well-annotated data, facilitating the early exploration of key covariates influencing recognition accuracy. However, as shown in Fig. 1(b), these datasets fall short regarding **population-level diversity**, as subjects of them contain the same limited group of covariates. Conversely, the emergence of outdoor datasets effectively addresses this

*Corresponding Author

Dataset	#Id	#Seq	#Cam	Data types	Covariates except view	Environment	Diversity
CMU MoBo	25	600	6	RGB, Sil.	TR, Speed, BA, IN	Controlled	Not Rich
SOTON	115	2,128	2	RGB, Sil.	TR	Controlled	Not Rich
USF	122	1,870	2	RGB	CO, GR, SH, BR, DU	Controlled	Not Rich
CASIA-B	124	13,640	11	RGB, Sil.	Coat, Bag	Controlled	Not Rich
CASIA-C	153	1,530	1	Inf., Sil.	SP, Bag	Controlled	Not Rich
OU-ISIR Speed	34	612	1	Sil.	TR, Speed	Controlled	Not Rich
OU-ISIR Cloth	68	2,764	1	Sil.	TR, CL	Controlled	Not Rich
OU-ISIR MV	168	4,200	25	Sil.	TR	Controlled	Not Rich
OU-LP	4,007	7,842	2	Sil.	None	Controlled	Not Rich
TUM GAID	305	3,370	1	RGB, Depth, A.	DU, BAC, SH	Controlled	Not Rich
OU-LP Age	63,846	63,846	1	Sil.	Age	Controlled	Not Rich
OU-MVLP	10,307	288,596	14	Sil., Pose, 3DM.	None	Controlled	Not Rich
OU-LP Bag	62,528	187,584	1	Sil.	Carrying	Controlled	Not Rich
GREW	26,345	128,671	882	Sil., Flow, Pose	Free walking	Wild	Population-Level
ReSGait	172	870	1	Sil., Pose	Free walking	Wild	Population-Level
UAV-Gait	202	9,895	6	Sil., pose	None	Controlled	Not Rich
Gait3D	4,000	25,309	39	Sil., Pose, 3DM.	Free walking	Wild	Population-Level
CASIA-E	1,014	778,752	26	RGB, Sil.	Bag, CL, WS	Controlled	Not Rich
CCPG	200	16,566	10	RGB, Sil.	CL	Controlled	Not Rich
SUSTech1K	1,050	25,279	12	RGB, Sil., 3DP	Bag, CL, UB, OC, NI	Controlled	Not Rich
CCGR (ours)	970	**1,580,617**	33	RGB, **Parsing**, Sil., Pose	**53 types per subject, as detailed in Figure 2.**	Controlled	**Population- and Individual-Level**

Table 1: Comparison of CCGR with existing datasets. Sil., Inf., A., and 3DM. mean silhouette, infrared, audio, and 3D Mesh&SMPL. #Id, #Seq, and #Cam refer to the number of identities, sequences, and cameras. BAC, CO, GR, BR, DU, IN, BA, TR, SH, CL, UB, OC, NI, and WS are abbreviations of backpack, concrete, grass, briefcase, duration, incline, ball, treadmill, shoes, clothing, umbrella, uniform, occlusion, night and walking style. CMU MoBo (Gross and Shi 2001); SOTON (Shutler et al. 2004); USF (Sarkar et al. 2005); CASIA-B (Yu, Tan, and Tan 2006); CASIA-C (Tan et al. 2006); OU-ISIR Speed (Mansur et al. 2014); OU-ISIR Cloth (Altab Hossain et al. 2010); OU-ISIR MV (Makihara, Mannami, and Yagi 2011); OU-LP (Iwama et al. 2012); TUM GAID (Hofmann et al. 2014); OU-LP Age (Xu et al. 2017); OU-MVLP (Takemura et al. 2018; An et al. 2020; Li et al. 2022); OU-LP Bag (Uddin et al. 2018); GREW (Zhu et al. 2021); ReSGait (Mu et al. 2021); UAV-Gait (Ding et al. 2022); Gait3D (Zheng et al. 2022); CASIA-E (Song et al. 2022); CCPG (Li et al. 2023); SUSTech1K (Shen et al. 2023).

limitation due to their real-world collection scenarios. Although their data distribution closely mirrors practical applications, we contend that current outdoor gait datasets lack **individual-level diversity**, as each subject typically contributes no more than seven variants (sequences) on average. This situation gives rise to two potential drawbacks for research: a) A majority of data pairs may qualify as "easy cases" owing to limited collection areas and short-term data gathering. b) The lack of fine annotations blocks exploring critical challenges relevant to real-world applications. More details of the existing dataset are in Table 1.

To overcome these limitations, we propose a novel gait recognition benchmark that introduces both population-level and individual-level diversity, named **Cross-Covariate Gait Recognition** or **CCGR**. Statistically, the CCGR dataset covers 970 subjects and approximately 1.6 million walking sequences. These sequences span 53 distinct walking conditions and 33 different filming views. Thus, each subject within CCGR ideally contains a comprehensive collection of $53 \times 33 = 1,749$ sequences. Notably, the walking conditions are widely distributed and well annotated, encompassing diverse factors such as carried items (book, bag, box, umbrella, trolley case, heavy bag, and heavy box), road types (up the stair, down the stair, up the ramp, down the ramp, bumpy road, soft road, and curved road), styles of walking (fast, stationary, normal, hands in pockets, free, and crowd), and more. The all-side camera array consisting of 33 cameras is installed at five different heights, effectively simulating the pitching angles of typical CCTVs. Every subject is recruited through a transparent process and accompanied by written consent. The age range of subjects spans from 6 to 70 years. The dataset encompasses raw RGB sequences. Releasing RGB images can facilitate the exploration of camera-based gait representations, and this paper officially provides common gait data like silhouette, parsing, and pose. CCGR will be made publicly available for research purposes.

Equipped with the proposed CCGR, we re-implement several representative state-of-the-art methods and investigate that: 1) Cross-covariate gait recognition is more challenging than that simulated by previous gait datasets, as the achieved best rank-1 accuracy is only 42.5%. 2) Certain less-researched covariates, such as the crowd, umbrella, overhead view, walking speed, road, mixed covariate, and more, significantly degrade the recognition accuracy. 3) The more covariates involved, regardless of population-level and individual-level diversity perspectives, the more challenging gait recognition becomes.

To solve complex covariate problems, this paper further introduces human parsing, which contains many semantic characteristics that describe body parts, to form a parsing-based baseline framework termed **ParsingGait**. In practice, we instantiate the backbone of PrasingGait using various

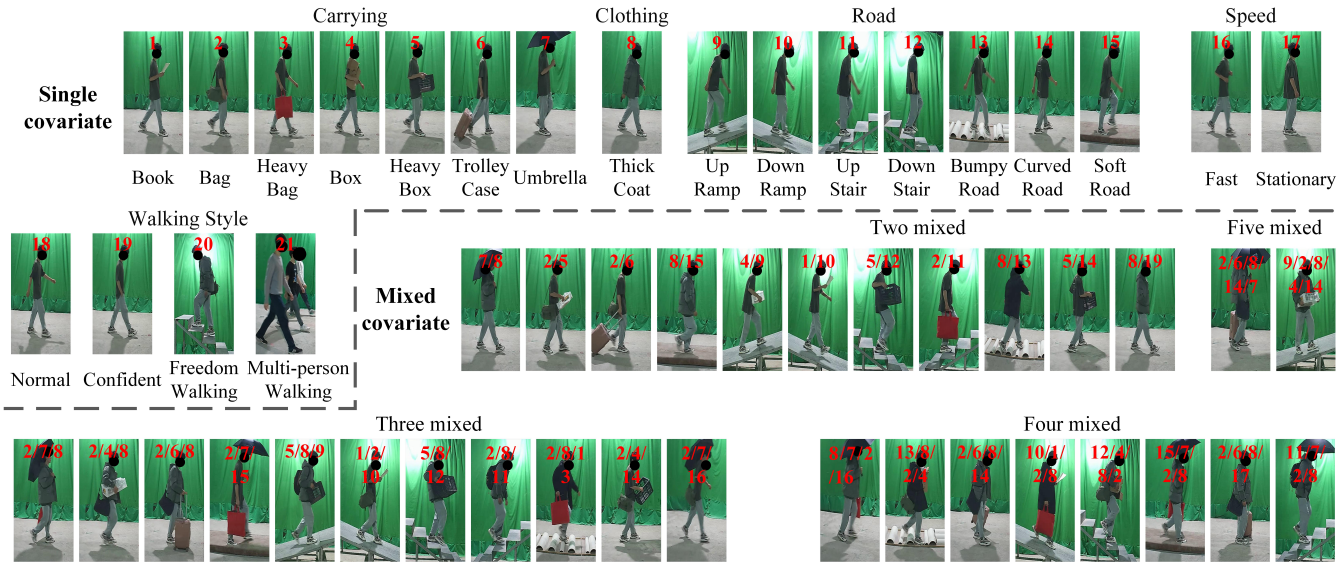


Figure 2: Examples of 53 covariates in CCGR. For a single covariate (the 1st row and the left of the 2nd row), the red numbers at the top of the pictures are indices of the covariates. For mixed covariates, numbers separated by "/" at the top of the picture indicate the co-occur of multi-single covariates corresponding to these numbers.

silhouette-based gait models, consistently achieving significant enhancements. By this means, this paper highlights the value of informative gait representations like human parsing images for gait pattern description.

In summary, our main contributions are as follows:

• We present the first well-annotated million-sequence-level gait recognition benchmark called CCGR, designed to research cross-covariate gait recognition deeply.

• We propose an efficient, compatible, and feasible parsing-based baseline framework named ParsingGait.

• We begin by evaluating existing algorithms to establish a baseline, then validating the effectiveness of ParsingGait. Next, we demonstrate the necessity of incorporating both population- and individual-level diversity. Finally, we thoroughly explore the impact of covariates and views.

The CCGR Benchmark

Covariates of CCGR

The dataset has 53 covariates; 21 are single covariates, while the remaining 32 are mixed covariates. Examples of the 53 covariates are shown in Figure 2.

Carrying: We have defined seven carrying covariates: **book**, **bag**, **heavy bag**, **box**, **heavy box**, and **trolley case**, **umbrella**. We have prepared 12 different types for the bag category, including single-shoulder bags, double-shoulder bags, satchels, backpacks, and handbags. Similarly, we have prepared eight boxes with varying shapes and volumes for the box category. As for the trolley case, we have prepared options in both 20-inch and 28-inch sizes. When subjects are asked to carry a bag, box, or trolley case, they can choose from the props we have provided. In the case of the heavy bag and box, we have placed counterweights inside them, ranging from 8kg to 15kg, to simulate the desired weight.

Clothing: Regarding the **thick coat** covariates, we have prepared a selection of 20 clothing items, which include down coats, overcoats, windbreakers, jackets, and cotton coats. When subjects are instructed to wear a thick coat, they can choose from our clothing collection.

Road: In addition to the normal road, we have prepared seven road covariates: **up/down the stair**, **up/down the ramp**, **bumpy road**, **soft (muddy) road**, and **curved road**. Ramps have a slope of 15^o. Curved road means subjects are asked to walk a curved track instead of a straight path.

Speed: In addition to the normal walking speed, we discuss two additional walking speeds: **fast** and **stationary**. Fast entails the subject walking at a speed close to a trot, while stationary refers to the subject remaining unmoving.

Walking Style: The remaining four single covariates include **normal walking**, **confident**, **multi-person walking**, and **freedom walking**. Normal walking indicates walking on a horizontal path at a normal speed without wearing a thick coat or carrying any items. Confident means that subjects place their hands inside their pant or clothing pockets. Multi-person walking means multiple subjects walking together. Freedom walking means subjects are free to choose their carrying, clothing, road, and speed.

Mixed covariates: In the real world, multiple covariates often co-occur. For instance, a man may wear a thick coat, carry a bag, and walk up a ramp. To simplify matters, we utilize mixed covariates to represent the co-occurrence of multiple covariates. In CCGR, we have designed 32 **mixed covariates** that are frequently encountered in daily life. Refer to Figure 2 for further details about these mixed covariates.

Views of CCGR

We rent a 500-square-meter warehouse and set up 33 cameras to collect data. Camera settings are shown in Figures

Figure 3: Examples of 33 views in CCGR. The red numbers at the top of the picture represent the horizontal angle.

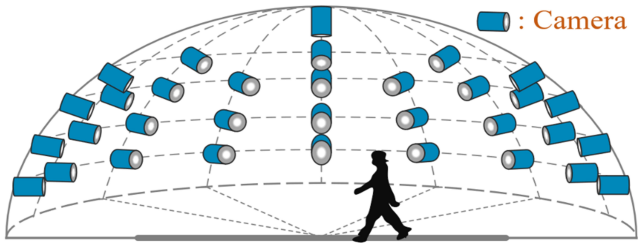


Figure 4: Camera setup in CCGR.

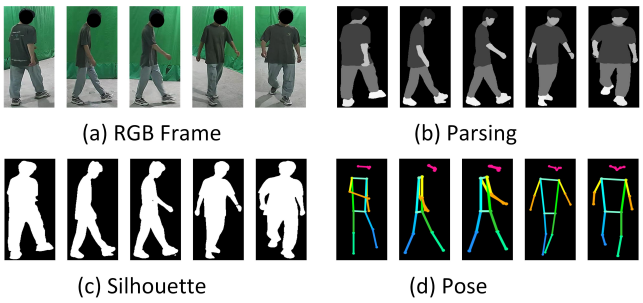


Figure 5: Examples of different gait data in CCGR .

4. The cameras are divided into five layers, from bottom to top. Layer 5 is the overhead camera with a pitch angle of 90^o. For the other four layers, the pitch angles from bottom to top are 5^o, 30^o, 55^o, and 75^o, and the horizontal angles of each layer increase from 0^o to 180^o counterclockwise. The frame size of the video files is 1280×720, and the frame rate is 25 fps. Figure 3 shows the example with various views.

Extraction of Multiple Gait Data

We offer various types of gait data, including RGB, parsing, silhouette, and pose; examples can be seen in Figure 5.

Parsing: Predicting the semantic category of each pixel on the human body is a fundamental task in computer vision, often referred to as human parsing (Liang et al. 2018; Zhao et al. 2018; Gong et al. 2018; Xia et al. 2017). We uses QANet (Yang et al. 2021) for parsing extraction. QANet takes an RGB image as its input and produces the semantic category of each pixel on the human body, including hair, face, and left leg. Initially, QANet employed integers ranging from 0 to 19 to represent these different categories. To facilitate visualization and image pruning, we multiply these integers by 13 to generate a grayscale image.

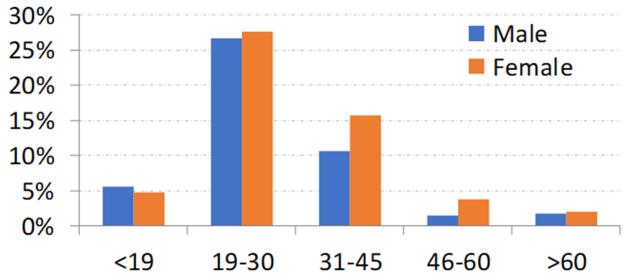


Figure 6: Age and gender attributes. Ages are categorized into five groups (< 19, $19-30$, $31-45$, $46-60$, and > 60).

Silhouette: We generate the silhouettes by directly binarizing the previously acquired parsing images. We have also tried the instance and semantic segmentation algorithms but attained relatively inferior gait recognition accuracy.

Pose: We use HRNet (Sun et al. 2019) to extract 2D Pose. We also try AlphaPose (Fang et al. 2017) and Openpose (Cao et al. 2017), which result in inferior accuracy.

Collection, Statistics and Evaluation

Collection Process: To simplify the description, we refer to covariates mentioned in the previous subsection as the "walking conditions". In the normal walking condition, each subject walks twice. In the remaining 52 walking conditions, each subject only walks once per condition. Therefore, a total of 54 walks per subject are required. Since each subject has to walk 54 times, and the walking conditions have to be changed each time, it takes 2 hours to collect one subject.

Dataset statistics: Figure 6 presents the distribution of age and gender in CCGR. The proportions of the various covariates align with the number of walks for each covariate. Furthermore, CCGR exhibits an average of 110 frames per sequence, more than 94% of sequences with > 60 frames.

Evaluation Protocol: Subjects are labeled from 1 to 1000. Subjects 134 to 164 are missing. Subjects 1 to 600 are used for training, and the rest are used for testing. The

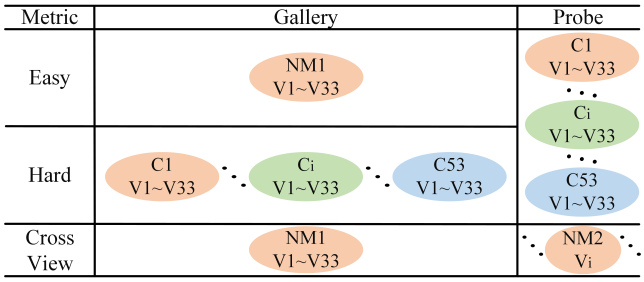


Figure 7: Evaluation metrics. C and V denote covariates and views, where subscripts indicate the order. NM is normal walking. "Easy" is employed by CASIA-B and OUMVLP (the gallery is normal walking). "Hard" is similar to GREW and Gait3D, closer to real life (the gallery is uncertain).

evaluation metrics are illustrated in Figure 7.

Parsing-based Gait Recognition

Although silhouette and pose are commonly employed as gait modalities, they possess significant limitations. Silhouette provides only contour information, while pose offers solely structural details, resulting in sparse and simplistic representations. Consequently, these modalities prove less effective when confronted with complex covariate environments. We are fortunate to discover that parsing can simultaneously provide contour, structural and semantic information. Notably, parsing eliminates texture and color, providing a basis for treating as a gait pattern.

Parsing and silhouettes have similar data structures, enabling parsing to inherit all silhouette-based algorithms without modification. This convenient compatibility allows us to explore parsing-based gait recognition efficiently. This paper explores the effectiveness of "Parsing + silhouette-based algorithms" and calls it **ParsingGait**.

Baseline on CCGR

Appearance-based Approaches

We evaluate some SOTA algorithms: GEINet (Shiraga et al. 2016), GaitSet (Chao et al. 2019), GaitPart (Fan et al. 2020), CSTL (Huang et al. 2021), GaitGL (Lin, Zhang, and Yu 2021), GaitBase (Fan et al. 2023b), and DeepGaitV2 (Fan et al. 2023a). **Implementation details:** All silhouettes are aligned by the approach mentioned in (Takemura et al. 2018) and transformed to 64×44. The batch size is $8 \times 16 \times 30$, where 8 denotes the number of subjects, 16 denotes the number of training samples per subject, and 30 is the number of frames. The optimizer is Adam. The number of iterations is 320K. The learning rate starts at 1e-4 and drops to 1e-5 after 200K iterations. For GaitBase and DeepGaitV2: The optimizer is SGD. The number of iterations is 240K. The learning rate starts at 1e-1 and drops by $1/10$ at 100k, 140k, and 170k. All models are trained on the entire training set.

Model-based Approaches

We evaluate two SOTA algorithms: GaitGraph (Teepe et al. 2021) and GaitGraph2 (Teepe et al. 2022). We train GaitGraph for 1200 epochs with a batch size of 128. We train GaitGraph2 for 500 epochs with a batch size of 768.

Methods	**R-1**^{hard}	R-1^{easy}	R-5^{easy}	R-5^{hard}
GEINet	3.10	4.62	9.20	12.7
GaitSet	25.3	35.3	46.7	58.9
GaitPart	22.6	32.7	42.9	55.5
GaitGL	23.1	35.2	39.9	54.1
CSTL	7.25	11.8	13.79	20.1
GaitBase	31.3	43.8	51.3	64.4
DeepGaitV2	42.5	55.2	63.2	75.2
GaitGraph	15.2	25.2	37.2	51.6
GaitGraph2	0.26	0.27	1.4	1.41

Table 2: The accuracy of representative methods on CCGR.

Backbone	**R-1**^{hard}	R-1^{easy}	R-5^{hard}	R-5^{easy}
GaitSet	31.6	42.8	54.8	67
GaitPart	29.0	40.9	51.5	64.5
GaitGL	28.4	42.1	46.6	61.4
CSTL	27.9	40.7	47.1	61.5
GaitBase	43.2	56.9	63.7	76.0
DeepGaitV2	52.7	67.2	74.7	87.7

Table 3: The accuracy of ParsingGait (ours) on CCGR.

Experiment

Analysis of Representative Methods

The results are shown in Table 2. The R-1^{hard} of GEINet, GaitSet, GaitPart, GaitGL, and CSTL falls below 26%. While these methods demonstrate near 90% accuracy on previous indoor datasets, their validity under complex covariates remains untested. On the other hand, GaitGraph and GaitGraph2 exhibit poorer performance compared to silhouette-based methods, potentially because the pose information can be sparser than the silhouette, resulting in less available information.

GaitBase and DeepGaitV2 are proposed to address the challenge of outdoor datasets; they are more robust against complex covariates. However, **DeepGaitV2 achieves an impressive 82% rank-1 accuracy on the outdoor dataset GREW. In contrast, its performance on CCGR falls considerably below, reaching a mere 43%**. This disparity may be due to the lack of individual-level diversity in the existing outdoor datasets.

Analysis of Parsing-based Gait Recognition

As shown in Table 3, the accuracy of ParsingGait is substantially improved. These findings effectively illustrate the three main advantages of parsing: feasibility, validity, and compatibility. By distinguishing between different body parts, parsing makes it more robust in the face of complex covariates. ParsingGait is the same computationally efficient as its silhouette-based counterpart because our parsing is consistent with the silhouette data structure.

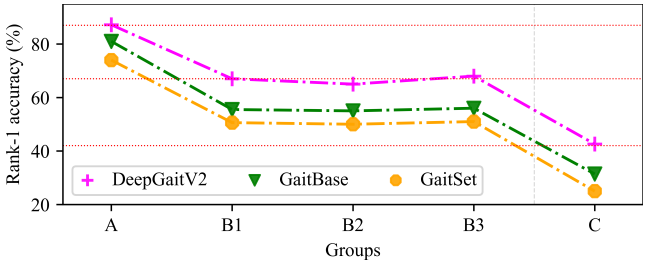


Figure 8: Increasing population and individual diversity.

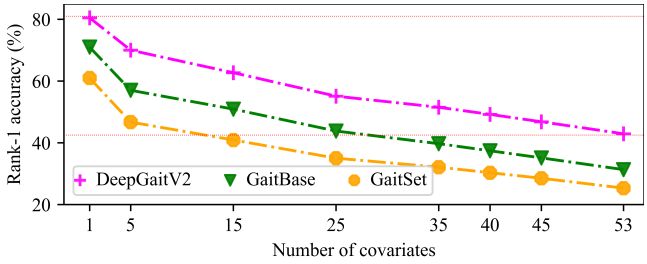


Figure 9: Impact of the number of covariates.

	Similar	Sample Setup	NoC per Sbj	NoC in Sub-dataset
A	CASIAB	NM BG CL L1	3	3
B1	Gait3D/ GREW	Random 8 Seqs per Sbj	Max 8	53
B2			Max 8	53
B3			Max 8	53
C	Ours	All Seqs	5	53

Table 4: Covariate Sampling Setup: L1, Seq, Sbj, and NoC refer to layer 1, sequence, subject, and number of covariates.

Population and Individual-Level Diversity

We research the impact of covariate diversity by sampling and isolating various covariates. The specific sampling setup is provided in Table 4. The experiments are categorized into five groups. Group A represents the absence of covariate diversity, while B1, B2, and B3 demonstrate population-level diversity without individual-level diversity. Lastly, Group C exhibits both population-level and individual-level diversity.

Based on the experimental data in Figure 8. From A to B1/2/3, the accuracy averagely decreased by -18.6%. However, from B1/2/3 to C, the accuracy averagely decreased by -25.1%. These findings indicate that **relying solely on population-level diversity is insufficient to accurately represent the underlying challenge, while individual-level diversity also is a significant challenge**. In addition, the trend of Figure 8 is generally consistent with Figure 1 at the beginning of the paper, further strengthening the credibility of the experimental results.

Impact of the Number of Covariates

We examine how the number of covariates impacts accuracy, and the experimental outcomes are illustrated in Figure 9. The accuracy is substantially decreased as we progressively increase the covariate number from 1 to 53. Furthermore, a troubling trend emerges: even when the number reaches 53, the decline in accuracy rate does not significantly decelerate. This observation may indicate that gait recognition faces greater challenges in real-world scenarios.

Evaluation of Covariates and Views

Single-Covariate Evaluation: As shown in Table 5. Multi-person walking significantly affects accuracy because many parts of the human body are obscured. Speed also significantly affects the accuracy as it dramatically impacts the temporal feature extraction of the algorithm. Clothing is still a big challenge. In addition, carrying and road also have a notable negative impact on accuracy.

Mixed-Covariate Evaluation: As shown in Table 6. Mixed covariates impact precision more, with a significant classical decrease as the number of mixes increases. for example, "Bag → BG-TC → BG-TC-CL → BG-TC-CL-ST", accuracy is gradually declining. However, mixed covariates are a challenge that must be addressed because ideal condi-

Gallery: Normal 1				
Type	Covariate	Gait Base	Deep GaitV2	Parsi ngGait
Carry	Book(BK)	65.7	75.3	85.5
	Bag(BG)	64.9	75.4	86.1
	Heavy Bag(HVBG)	60.0↓	72.3	84.2
	Box(BX)	61.5	71.6	83.0
	Heavy Box(HVBX)	58.7↓	69.7↓	81.9↓
	Trolley Case(TC)	64.1	73.0	83.4
	Umbrella(UB)	47.2↓	60.5↓	71.3↓
	Average	60.3	71.1	**82.2**
Cloth	Thick Coat(CL)	40.4	53.5	66.8
Road	Up Ramp(UTR)	60.3↓	69.5↓	80.9
	Down Ramp(DTR)	60.5↓	70.1↓	80.2
	Up Stair(UTS)	54.9↓	66.7↓	78.0↓
	Down Stair(DTS)	54.0↓	65.4↓	76.7↓
	Bumpy Road(BM)	63.3	71.4	82.0
	Curved Road(CV)	70.0	77.3	86.1
	Soft Road(SF)	66.0	73.2	83.7
	Average	61.3	70.5	**79.3**
Speed	Normal 1(NM1)	76.6	83.5	91.3
	Fast(FA)	47.2↓	60.7↓	74.1↓
	Stationary(ST)	32.0↓	45.0↓	60.9↓
	Average	51.9	63.1	**75.4**
Walk Style	Normal 2(NM2)	75.3	82.3	90.7
	Confident(CF)	64.9	74.8	83.9
	Freedom(FD)	57.1	68.1	79.2
	Multi-person(MP)	24.0↓	32.6↓	39.4↓
	Average	55.3	64.4	**73.3**

Table 5: Single-Covariate Evaluation: R-1^{easy} accuracy (%) with excluding identical-view cases. ↓ and *bold* respectively indicate the sub-*Average* and SOTA performance.

Gallery: Normal 1				
Type	Covariate	Gait Base	Deep GaitV2	Parsi ngGait
Two Mixed	CL-UB	25.2↓	37.8↓	46.9↓
	HVBX-BG	52.1	64.7	78.3
	BG-TC	58.1	69.3	81.3
	SF-CL	36.1↓	48.0↓	62.8↓
	UTR-BX	51.0	62.0	75.4
	DTR-BK	55.1	66.0	77.4
	DTS-HVBX	42.6↓	56.1↓	69.8↓
	UTS-BG	46.8	60.9	74.5
	BM-CL	35.2↓	46.3	61.8
	CV-HVBX	61.0	70.8	82.0
	CL-CF	39.2↓	52.7↓	65.6↓
	Average	45.7	57.7	**70.5**
Three Mixed	CL-UB-BG	23.4↓	36.1↓	44.9↓
	BX-BG-CL	35.1↓	48.8	60.7
	BG-TC-CL	34.3↓	48.5	63.0
	SF-UB-BG	36.4↓	49.4	62.5
	UTR-HVBX-CL	31.8↓	43.1↓	55.3↓
	DTR-BK-BG	49.2	61.7	74.9
	DTS-HVBX-CL	26.4↓	38.0↓	49.1↓
	UTS-BG-CL	25.1↓	37.7↓	52.5↓
	BM-CL-BG	33.0↓	44.8↓	59.6↓
	CV-BX-BG	58.8	69.6	80.8
	UB-BG-FA	28.0↓	41.0↓	52.8↓
	Average	34.7	47.1	**59.7**
Four Mixed	CL-UB-BG-FA	16.2↓	27.6↓	35.7↓
	BM-CL-BG-BX	32.2	43.5	56.1
	BG-TC-CL-CV	38.0	51.2	66.9
	DTR-BK-BG-CL	32.2	44.9	56.9
	DTS-BX-CL-BG	25.6	37.3	48.9
	SF-UB-BG-CL	20.6↓	31.8↓	41.9↓
	BG-TC-CL-ST	11.7↓	18.4↓	29.4↓
	UTS-UB-BG-CL	15.8↓	26.1↓	36.4↓
	Average	24.0	35.1	**46.5**
Five Mixed	BG-TC-CL-CV-UB	34.1	35.9	**47.4**
	UTR-BG-CL-BX-CV	31.3	45.2	**58.3**

Table 6: Mixed-Covariate Evaluation: R-1^{easy} accuracy (%) with excluding identical-view cases. We use "-" to connect the mixed covariates. Tab. 5 presents the dictionary containing abbreviations and their corresponding full spellings of these covariates. ↓ and *bold* respectively indicate the sub-*Average* and SOTA performance.

Cross-view Evaluation				
Pitch Angle	Probe View	Gait Base	Deep GaitV2	Parsi ngGait
5°	0.0°	80.1	85.7	90.6
	22.5°	84.7	89.5	93.1
	45.0°	83.7	89.1	93.9
	67.5°	79.3	85.7	93.6
	90.0°	75.7	83.7	93.2
	112.5°	76.9	84.6	93.2
	135.0°	81.6	87.1	93.7
	157.5°	83.8	88.6	92.7
	180.0°	77.4	83.3	89.9
	Average	80.4	86.4	92.6
30°	0.0°	79.6	85.2	92.0
	22.5°	85.0	89.8	93.6
	45.0°	86.0	90.9	94.9
	67.5°	82.7	88.8	95.0
	90.0°	78.9	86.4	94.6
	112.5°	79.1	86.3	94.5
	135.0°	82.8	88.5	94.5
	157.5°	84.1	89.9	93.7
	180.0°	79.5	85.3	91.8
	Average	82.0	87.9	93.9
55°	0.0°	74.8	81.8	90.6
	22.5°	81.5	86.7	93.3
	45.0°	83.9	88.9	95.0
	67.5°	82.2	88.4	95.1
	90.0°	63.6	76.3	92.0
	112.5°	77.3	84.5	93.6
	135.0°	81.2	87.4	93.9
	157.5°	80.8	86.3	93.2
	180.0°	75.9	83.2	91.3
	Average	77.9	84.8	93.1
75°	0.0°	64.4	74.8	86.0
	45.0°	78.7	85.2	92.7
	90.0°	40.8	60.9	87.5
	135.0°	73.2	80.5	90.6
	180.0°	62.5	74.0	86.2
	Average	63.9	75.1	88.6
OverHead	-	2.0	8.4	32.0

Table 7: Cross-View Evaluation: Rank-1 accuracy (%) with excluding identical-view cases.

tions for single covariates in real life tend to be rare.

Cross-View Evaluation: As shown in Table 7. The existing algorithms perform well, considering only the views. The current challenge with views is how to address the high-pitch angle case. Encouragingly, ParsingGait demonstrates distinct improvement in recognizing overhead views.

Conclusion

This paper introduces CCGR, a well-labeled dataset which provides diversity at both the population and individual levels. As gait recognition on many public gait datasets is close to saturation, future works can explore how gait is affected by covariates and how to design robust gait recognition.

Acknowledgements

This work wad supported by the Natural Science Foundation of Hunan Province (No.2023JJ30697), the Changsha Natural Science Foundation (No.kq2208286) and the National Natural Science Foundation of China (No.61502537). This work was also supported in part by the National Key Research, and in part by Development Program of China under Grant (No.61976144) and the Shenzhen International Research Cooperation Project under Grant (No.GJHZ20220913142611021).

References

Altab Hossain, M.; Makihara, Y.; Wang, J.; and Yagi, Y. 2010. Clothing-invariant gait identification using part-based clothing categorization and adaptive weight control. *PR*, 43(6): 2281–2291.

An, W.; Yu, S.; Makihara, Y.; Wu, X.; Xu, C.; Yu, Y.; Liao, R.; and Yagi, Y. 2020. Performance Evaluation of Model-based Gait on Multi-view Very Large Population Database with Pose Sequences. *IEEE Trans. on Biometrics, Behavior, and Identity Science*.

Cao, Z.; Simon, T.; Wei, S.-E.; and Sheikh, Y. 2017. Real-time Multi-Person 2D Pose Estimation Using Part Affinity Fields. In *CVPR*.

Chao, H.; He, Y.; Zhang, J.; and Feng, J. 2019. GaitSet: Regarding Gait as a Set for Cross-View Gait Recognition. In *AAAI*.

Ding, T.; Zhao, Q.; Liu, F.; Zhang, H.; and Peng, P. 2022. A Dataset and Method for Gait Recognition with Unmanned Aerial Vehicless. In *ICME*.

Fan, C.; Hou, S.; Huang, Y.; and Yu, S. 2023a. Exploring Deep Models for Practical Gait Recognition. *ArXiv*, abs/2303.03301.

Fan, C.; Liang, J.; Shen, C.; Hou, S.; Huang, Y.; and Yu, S. 2023b. OpenGait: Revisiting Gait Recognition Towards Better Practicality. In *CVPR*, 9707–9716.

Fan, C.; Peng, Y.; Cao, C.; Liu, X.; Hou, S.; Chi, J.; Huang, Y.; Li, Q.; and He, Z. 2020. GaitPart: Temporal Part-Based Model for Gait Recognition. In *CVPR*.

Fang, H.-S.; Xie, S.; Tai, Y.-W.; and Lu, C. 2017. RMPE: Regional Multi-Person Pose Estimation. In *ICCV*.

Gong, K.; Liang, X.; Li, Y.; Chen, Y.; Yang, M.; and Lin, L. 2018. Instance-Level Human Parsing via Part Grouping Network. In *ECCV*, 805–822. ISBN 978-3-030-01225-0.

Gross, R.; and Shi, J. 2001. The CMU Motion of Body (MoBo) Database. *Monumenta Nipponica*.

Hofmann, M.; Geiger, J.; Bachmann, S.; Schuller, B.; and Rigoll, G. 2014. The TUM Gait from Audio, Image and Depth (GAID) database: Multimodal recognition of subjects and traits. *JVCIR*, 25(1): 195–206.

Huang, X.; Zhu, D.; Wang, H.; Wang, X.; Yang, B.; He, B.; Liu, W.; and Feng, B. 2021. Context-Sensitive Temporal Feature Learning for Gait Recognition. In *ICCV*, 12909–12918.

Iwama, H.; Okumura, M.; Makihara, Y.; and Yagi, Y. 2012. The OU-ISIR Gait Database Comprising the Large Population Dataset and Performance Evaluation of Gait Recognition. *IEEE Trans. on Information Forensics and Security*, 7, Issue 5: 1511–1521.

Li, W.; Hou, S.; Zhang, C.; Cao, C.; Liu, X.; Huang, Y.; and Zhao, Y. 2023. An In-Depth Exploration of Person Re-Identification and Gait Recognition in Cloth-Changing Conditions. In *CVPR*, 13824–13833.

Li, X.; Makihara, Y.; Xu, C.; and Yagi, Y. 2022. Multi-View Large Population Gait Database With Human Meshes and Its Performance Evaluation. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 4(2): 234–248.

Liang, X.; Gong, K.; Shen, X.; and Lin, L. 2018. Look into Person: Joint Body Parsing & Pose Estimation Network and a New Benchmark. *IEEE TPAMI*.

Lin, B.; Zhang, S.; and Yu, X. 2021. Gait Recognition via Effective Global-Local Feature Representation and Local Temporal Aggregation. In *ICCV*, 14648–14656.

Makihara, Y.; Mannami, H.; and Yagi, Y. 2011. Gait Analysis of Gender and Age Using a Large-Scale Multi-view Gait Database. In Kimmel, R.; Klette, R.; and Sugimoto, A., eds., *ACCV*, 440–451. Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN 978-3-642-19309-5.

Mansur, A.; Makihara, Y.; Aqmar, R.; and Yagi, Y. 2014. Gait Recognition under Speed Transition. In *CVPR*.

Mu, Z.; Castro, F. M.; Marín-Jiménez, M. J.; Guil, N.; ran Li, Y.; and Yu, S. 2021. ReSGait: The Real-Scene Gait Dataset. In *IJCB 2021*.

Sarkar, S.; Phillips, P.; Liu, Z.; Vega, I.; Grother, P.; and Bowyer, K. 2005. The humanID gait challenge problem: data sets, performance, and analysis. *IEEE TPAMI*, 27(2): 162–177.

Shen, C.; Fan, C.; Wu, W.; Wang, R.; Huang, G. Q.; and Yu, S. 2023. LidarGait: Benchmarking 3D Gait Recognition With Point Clouds. In *CVPR*, 1054–1063.

Shiraga, K.; Makihara, Y.; Muramatsu, D.; Echigo, T.; and Yagi, Y. 2016. GEINet: View-invariant gait recognition using a convolutional neural network. In *ICB*, 1–8.

Shutler, J. D.; Grant, M. G.; Nixon, M. S.; and Carter, J. N. 2004. On a large sequence-based human gaitdatabase. In *Applications and Science in Soft Computing*.

Song, C.; Huang, Y.; Wang, W.; and Wang, L. 2022. CASIA-E: a large comprehensive dataset for gait recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(3): 2801–2815.

Sun, K.; Xiao, B.; Liu, D.; and Wang, J. 2019. Deep High-Resolution Representation Learning for Human Pose Estimation. In *CVPR*.

Takemura, N.; Makihara, Y.; Muramatsu, D.; Echigo, T.; and Yagi, Y. 2018. Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition. *IPSJ Transactions on Computer Vision and Applications*, 10.

Tan, D.; Huang, K.; Yu, S.; and Tan, T. 2006. Efficient Night Gait Recognition Based on Template Matching. In *ICPR)*, volume 3, 1000–1003.

Teepe, T.; Gilg, J.; Herzog, F.; Hörmann, S.; and Rigoll, G. 2022. Towards a Deeper Understanding of Skeleton-Based Gait Recognition. In *CVPRW*.

Teepe, T.; Khan, A.; Gilg, J.; Herzog, F.; Hörmann, S.; and Rigoll, G. 2021. Gaitgraph: Graph Convolutional Network for Skeleton-Based Gait Recognition. In *2021 IEEE International Conference on Image Processing (ICIP)*, 2314–2318.

Uddin, M. Z.; Ngo, T. T.; Makihara, Y.; Takemura, N.; Li, X.; Muramatsu, D.; and Yagi, Y. 2018. The OU-ISIR Large Population Gait Database with real-life carried object and its performance evaluation. *IPSJ Transactions on Computer Vision and Applications*, 10(1): 5.

Xia, F.; Wang, P.; Chen, X.; and Yuille, A. L. 2017. Joint Multi-Person Pose Estimation and Semantic Part Segmentation. In *CVPR*.

Xu, C.; Makihara, Y.; Ogi, G.; Li, X.; Yagi, Y.; and Lu, J. 2017. The OU-ISIR Gait Database Comprising the Large Population Dataset with Age and Performance Evaluation of Age Estimation. *IPSJ Trans. on Computer Vision and Applications*, 9(24): 1–14.

Yang, L.; Song, Q.; Wang, Z.; Liu, Z.; Xu, S.; and Li, Z. 2021. Quality-Aware Network for Human Parsing. In *arXiv preprint arXiv:2103.05997*.

Yu, S.; Tan, D.; and Tan, T. 2006. A Framework for Evaluating the Effect of View Angle, Clothing and Carrying Condition on Gait Recognition. In *ICPR*, volume 4, 441–444.

Zhao, J.; Li, J.; Cheng, Y.; Sim, T.; Yan, S.; and Feng, J. 2018. Understanding Humans in Crowded Scenes: Deep Nested Adversarial Learning and A New Benchmark for Multi-Human Parsing. In *ACM MM*, 792–800. ISBN 9781450356657.

Zheng, J.; Liu, X.; Liu, W.; He, L.; Yan, C.; and Mei, T. 2022. Gait Recognition in the Wild With Dense 3D Representations and a Benchmark. In *CVPR*, 20228–20237.

Zhu, Z.; Guo, X.; Yang, T.; Huang, J.; Deng, J.; Huang, G.; Du, D.; Lu, J.; and Zhou, J. 2021. Gait Recognition in the Wild: A Benchmark. In *ICCV*, 14789–14799.