Supplementary Material for Cut out the Middleman: Revisiting Pose-based Gait Recognition

Yang Fu[1], Saihui Hou[1,2(✉)], Shibei Meng[1], Xuecai Hu[1(✉)], Chunshui Cao[2], Xu Liu[2], and Yongzhen Huang[1,2]

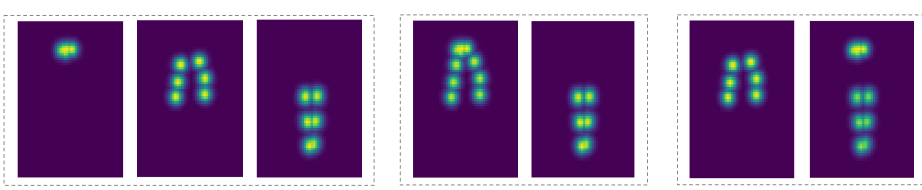
[1] School of Artificial Intelligence, Beijing Normal University
[2] WATRIX.AI

6 More Experiments

Impact of Different Partition Strategies. The partition strategy significantly improves local feature extraction by focusing on detailed segments of the body, thereby enhancing the capture of fine-grained characteristics. By implementing various partitioning strategies, as illustrated in Figure 6, multiple channels are integrated, each representing unique segments of the human body, to maximize the utilization of semantic information across different body parts. As demonstrated in Table 6, the third partition strategy outperforms others in terms of performance, leading to its selection as our preferred strategy.

Table 6: Effect of different human partition strategies.

Method	CASIA-B				CCPG							
	NM	BG	CL	Mean	CL		UP		DN		BG	
	R-1 (%)				R-1 & mAP(%)							
3-Part	97.95	91.12	71.61	86.89	72.65	40.76	81.20	**58.05**	84.94	60.20	89.33	65.64
2-Part-A	97.68	90.76	72.60	87.01	69.77	37.66	80.94	54.41	83.06	56.37	87.54	62.65
2-Part-B	**98.22**	**92.14**	**74.15**	**88.17**	**73.59**	**41.06**	**83.19**	58.04	**86.47**	**60.27**	**90.53**	**66.36**



(a) 3-Part (b) 2-Part-A (c) 2-Part-B

Fig. 6: The visualization of three different partition strategies.

Table 7: Effect of various network architectures.

Method	inference embedding dimension	CASIA-B				CCPG							
		NM	BG	CL	Mean	CL		UP		DN		BG	
		R-1 (%)				R-1 & mAP(%)							
Glocal	16×256	97.43	89.19	73.99	86.87	68.79	36.64	80.42	53.12	83.92	55.56	86.95	60.84
Gloabl-Local	16×256×2	98.18	**93.06**	**74.87**	**88.70**	71.88	38.32	82.24	55.61	83.06	56.98	89.42	65.65
Globa-Local with Fusion	16×256	**98.22**	92.14	74.15	88.17	**73.59**	**41.06**	**83.19**	**58.04**	**86.47**	**60.27**	**90.53**	**66.36**

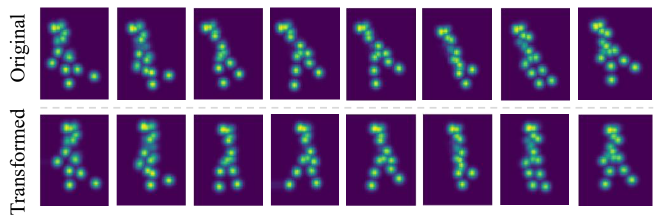


Fig. 7: The visualization of Pose-Guided Heatmap Alignment.

Analysis of Global-Local Network. As shown in Tab. 7, the performance of the global-local network surpasses that of the global branch alone, underscoring the significant contribution of the local branch to enhancing accuracy. When compared to the network featuring a fusion branch, incorporating a fusion branch not only facilitates a reduction in the gait representation's dimensionality, which can expedite the retrieval process and minimize the storage requirements for gait embeddings, but also maintains results comparable to those achieved with larger embeddings derived from two branches. This demonstrates the capacity of the fusion branch to balance efficiency and performance effectively.

Visualization of PGHA. In Fig. 7, we visualize some examples of the Heatmap Alignment, demonstrating that the proposed module effectively eliminates body tilt and bias caused by the camera viewpoint, thereby enhancing the model's performance.

Comparison with Additional Methods. In Tab. 8, we compared the results of methods based on NAS and other latest methods on CASIA-B. **(1)** Distinct from SPOSGait [3] and CLASH [2], our method is simple yet effective, not relying on architecture search while still demonstrating stable performance. **(2)** Compared

Table 8: Rank-1 comparison with recent methods on CASIA-B.

Method	Modality	NM	BG	CL	Mean
SPOSGait (Arxiv'22)	Silhouette	-	-	-	92.12
DyGait (ICCV'23)		98.40	96.20	87.80	94.13
DANet (CVPR'23)		98.00	95.90	89.90	94.60
CLASH (TIP'24)		98.30	95.30	88.00	93.90
GaitHeat (Ours)	Pose	98.22	92.14	74.15	88.17
GaitHeat++ (Ours)		**99.60**	**97.88**	**90.35**	**95.94**

with DyGait [7] and DANet [5], our method is based on human pose instead of silhouettes. Thanks to the shape information provided by the heatmap, our method achieves competitive results.

The Different View Results. Tab. 9 presents the comparative results across different views and clothing conditions compared to the previous pose-based methods. This is attributed to the generalized input representation of GaitHeat, where the continuous distribution of heatmap is more robust to keypoint prediction errors compared with the discrete coordinate used in previous work.

Table 9: Rank-1 on CASIA-B under all views and different conditions. We reproduce the multi-views results using OpenGait and FastPoseGait.

NM/BG/CL	0°-180°											
Method	0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°	Mean
GaitSet	90.80	97.90	99.40	96.90	93.60	91.70	95.00	97.80	98.90	96.80	85.80	95.00
GaitPart	94.10	98.60	99.30	98.50	94.00	92.30	95.90	98.40	99.20	97.80	90.40	96.20
GaitGL	96.00	98.30	99.00	97.90	96.90	95.40	97.00	98.90	99.30	98.80	94.00	97.40
GaitBase	95.30	99.50	100.00	99.10	97.00	95.60	98.00	99.40	**99.90**	99.20	93.80	97.89
GaitGraph	87.07	87.48	91.50	89.70	87.88	86.77	87.14	82.00	84.80	88.40	77.30	86.37
GaitGraph2	79.00	83.90	82.50	81.10	84.40	82.40	82.00	78.00	77.60	79.60	72.70	80.29
GaitTR	95.30	95.40	95.50	94.90	94.40	94.70	95.40	95.10	95.90	96.40	89.90	94.81
GPGait	87.10	96.10	97.10	96.70	92.30	91.00	93.50	96.50	97.70	96.20	85.30	93.60
GaitHeat	96.40	99.50	99.80	99.40	96.50	96.80	98.40	98.90	99.10	99.10	96.50	98.22
GaitHeat++	**99.60**	**99.90**	**100.00**	**100.00**	**99.30**	**99.40**	**99.60**	**99.70**	99.30	**99.80**	**99.00**	**99.60**
GaitSet	83.80	91.20	91.80	88.80	83.30	81.00	84.10	90.00	92.20	94.40	79.00	87.20
GaitPart	89.10	94.80	96.70	95.10	88.30	84.90	89.00	93.50	96.10	93.80	85.80	91.50
GaitGL	92.60	96.60	96.80	95.50	93.50	89.30	92.20	96.50	98.20	96.90	91.50	94.50
GaitBase	92.30	95.50	96.30	95.96	91.70	90.50	92.30	96.10	97.40	95.76	88.60	93.86
GaitGraph	81.31	80.60	82.73	76.77	74.29	76.34	75.31	73.94	76.00	76.60	67.60	76.50
GaitGraph2	73.00	77.60	72.50	73.50	71.80	70.80	71.70	70.40	71.40	74.10	58.60	71.40
GaitTR	86.40	89.10	89.00	89.10	88.50	87.40	87.10	88.50	90.30	89.90	78.80	87.65
GPGait	71.50	82.50	87.30	85.50	80.00	75.60	80.30	84.80	85.60	80.40	68.20	80.15
GaitHeat	90.60	94.60	96.00	95.20	90.40	88.00	90.90	92.20	95.30	92.10	88.20	92.14
GaitHeat++	**98.30**	**98.60**	**99.70**	**99.40**	**96.30**	**94.60**	**96.60**	**97.90**	**98.70**	**98.60**	**98.00**	**97.88**
GaitSet	61.40	75.40	80.70	77.30	72.10	70.10	71.50	73.50	73.50	68.40	50.00	70.40
GaitPart	70.70	85.50	86.90	83.30	77.10	72.50	76.90	82.20	83.80	80.20	66.50	78.70
GaitGL	76.60	90.00	90.30	87.10	84.50	79.00	84.10	87.00	87.30	84.40	69.50	83.60
GaitBase	69.40	80.20	82.70	81.30	76.70	75.00	76.10	78.90	81.00	78.90	66.90	77.01
GaitGraph	68.50	66.50	63.80	65.26	64.80	68.50	67.08	58.80	65.80	67.30	61.30	65.24
GaitGraph2	63.20	64.80	65.20	58.00	64.70	67.30	68.10	64.10	61.10	67.40	57.90	63.80
GaitTR	83.10	85.20	86.10	91.10	87.90	89.10	**91.90**	**91.00**	**91.00**	**90.60**	81.80	88.07
GPGait	58.10	72.80	76.00	73.30	67.30	66.10	71.00	75.80	72.60	71.70	57.60	69.29
GaitHeat	74.20	77.70	79.50	76.40	72.80	72.50	75.90	73.20	74.00	75.40	64.00	74.15
GaitHeat++	**90.10**	**91.90**	**92.00**	**92.70**	**92.40**	**91.80**	91.60	90.20	89.50	87.40	**84.30**	**90.35**

7 More Experiment Details

Details of the Datasets. We conducted experiments on three publicly available RGB gait datasets. **CASIA-B** [8] consists of 124 subjects, each subject under

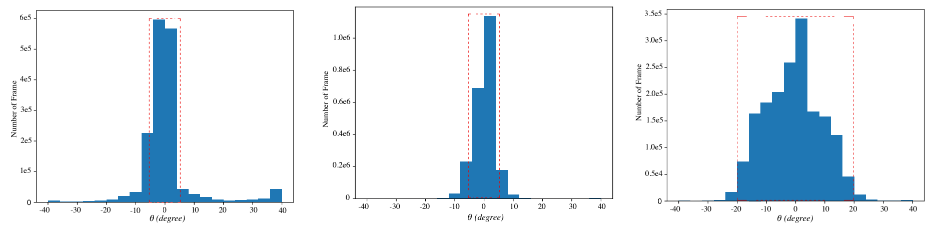
Table 10: The statistics of used datasets and implement details. Id and Seq present the number of identities and sequences, respectively. LR denotes the learning rate. The batch size $(8, 16)$ indicates sampling 8 subjects, with each subject comprising 16 sequences.

Dataset	Train Set		Test Set		Collection Situations	Batch Size	Optimizer	Scheduler		
	Id	Seq	Id	Seq				Type	Milestones	Total Steps
CASIA-B [8]	74	8140	50	5500	Indoor	(8,16)	SGD LR=0.1	Multi-step (drop by 1/10 per milestone)	[20k,40k,50k]	60k
CCPG [4]	100	8187	100	8178	Indoor & Outdoor	(8,16)			[20k,40k,50k]	60k
SUSTech1K [6]	200	6011	850	19228	Outdoor	(8,16)			[20k,40k,50k]	60k

three different statuses, which are normal, bagging, and clothing. The gait sequences are captured by 11 cameras at various angles, offering a comprehensive range of views. **CCPG** [4] includes 16K sequences from 200 subjects captured from indoor and outdoor scenes. It captures a broad range of clothing variations, with distinct cloth-changing statuses for each identity, including changes to the upper body, lower body, and full body. **SUSTech1K** [6] is notable for its wide range of attributes, containing common variations such as appearance, baggage, clothing, and different perspectives, as well as outdoor-specific challenges like occlusion, varied lighting conditions, uniform, and umbrella. It is a synchronized multimodal dataset, ensuring that frames across all modalities are timestamped consistently.

Implementation Details of Training. The specific hyper-parameters of our model are elaborated in Table 10. The batch sizes are set as (8, 16) across three benchmarks. We employ an SGD optimizer with a momentum of 0.9, coupled with a multi-step learning rate schedule. Triplet loss is applied with a margin of 0.2. The part number in Horizontal Pyramid Mapping (HPM) is set to 16.

Implementation Details of PGHA. Given the absence of the corresponding channels, we determine the positions for the neck P_{neck} and hip P_{hip} by calculating the average maximum response positions of the right and left shoulders for P_{neck}, and the right and left hips for P_{hip}, respectively. Additionally, the alignment threshold γ, as defined in PGHA, is set as 5° at CASIA-B and SUSTech1K, 20° for CCPG, based on our observation of the statistics of three datasets. As illustrated in Figure 8a, 8b, and 8c, the majority of rotation angles fall within the range of -5° to 5° at CASIA-B and SUSTech1K, and -20° to 20° for CCPG, which we consider as the normal human posture in the corresponding scenarios. Therefore, alignment is applied when the rotation angle θ exceeds the normal range threshold γ.



(a) The statistics of the rotation angles in CASIA-B.

(b) The statistics of the rotation angles in SUSTech1K.

(c) The statistics of the rotation angles in CCPG.

Fig. 8: The statistics of the rotation angles on three datasets.

8 Ethical Statements

Our work adheres to stringent ethical and security guidelines [1] of biometric, aiming to advance gait recognition technology for societal good and the enhancement of human welfare.

References

1. ISO/IEC 19795-1, information technology – biometric performance testing and reporting – part 1: Principles and framework, international organization for standardization std. (2006)
2. Dou, H., Zhang, P., Zhao, Y., Jin, L., Li, X.: Clash: Complementary learning with neural architecture search for gait recognition. IEEE Transactions on Image Processing (2024)
3. Guo, X., Zhu, Z., Yang, T., Lin, B., Huang, J., Deng, J., Huang, G., Zhou, J., Lu, J.: Gait recognition in the wild: A large-scale benchmark and nas-based baseline. arXiv e-prints pp. arXiv–2205 (2022)
4. Li, W., Hou, S., Zhang, C., Cao, C., Liu, X., Huang, Y., Zhao, Y.: An in-depth exploration of person re-identification and gait recognition in cloth-changing conditions. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 13824–13833 (2023)
5. Ma, K., Fu, Y., Zheng, D., Cao, C., Hu, X., Huang, Y.: Dynamic aggregated network for gait recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 22076–22085 (2023)
6. Shen, C., Fan, C., Wu, W., Wang, R., Huang, G.Q., Yu, S.: Lidargait: Benchmarking 3d gait recognition with point clouds. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1054–1063 (2023)
7. Wang, M., Guo, X., Lin, B., Yang, T., Zhu, Z., Li, L., Zhang, S., Yu, X.: Dygait: Exploiting dynamic representations for high-performance gait recognition. arXiv preprint arXiv:2303.14953 (2023)
8. Yu, S., Tan, D., Tan, T.: A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In: 18th international conference on pattern recognition (ICPR'06). vol. 4, pp. 441–444. IEEE (2006)