CarGait: Cross-Attention based Re-ranking for Gait recognition

Supplementary Material

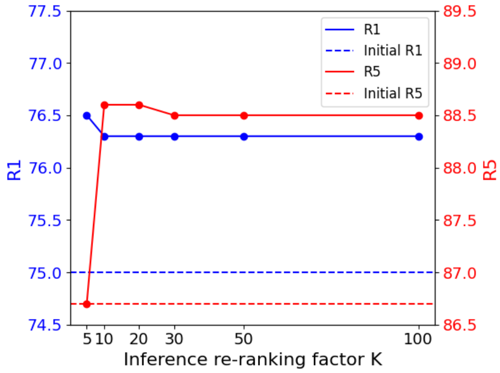

Figure 1. Ablation study on CarGait inference re-ranking factor K, with `SwinGait-3D` model trained on *Gait3D* dataset. Rank-1 (R1) and Rank-5 (R5) results are shown in blue (left-hand side) and red (right-hand side), respectively. The dashed lines indicate the initial single-stage model performance, while the solid lines represent the results after CarGait re-ranking.

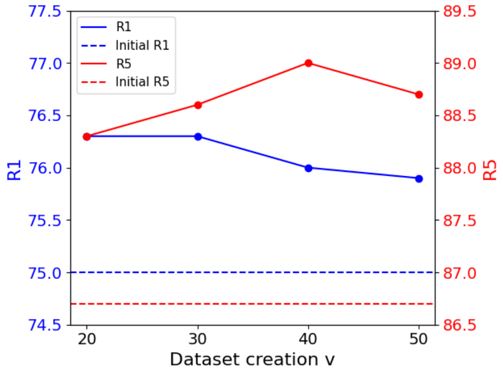

Figure 2. Ablation study on CarGait training dataset creation v, with `SwinGait-3D` model trained on *Gait3D* dataset. Rank-1 (R1) and Rank-5 (R5) results are shown in blue (left-hand side) and red (right-hand side), respectively. The dashed lines indicate the initial single-stage model performance, while the solid lines represent the results after CarGait re-ranking.

1. Hyperparameters

In the paper, we present an ablation study on the hyperparameters used in CarGait, K and v. Both are fixed across all the experiments shown in the paper. The inference re-ranking factor K, which is the length of the top-ranked list on which re-ranking is performed during inference, is set to 10. The size of the candidate set in training, v, is set to 30. Here, we present a detailed analysis on K and v, using the `SwinGait-3D` model [3] trained on *Gait3D* dataset [11].

As shown in Fig. 1, CarGait improvements are consistent across different values of K, with minor changes in R1 (higher for $K = 5$), and in R5 (higher for $K = 10$ and $K = 20$). Figure 2 presents an ablation study on the top-v candidates (per probe) used to construct the training set, as illustrated in the paper (Method section). Compared to the inference factor K, the performance variations here are more pronounced. Nevertheless, for all values of v shown (solid lines in Fig. 2), both Rank-1 and 5 accuracy consistently surpass the initial state (dashed lines). As mentioned, CarGait with $v = 30$ has been selected as a fixed hyperparameter for all `models` and *datasets* to ensure better generalization of our method. That is, even though, in some cases, it might not be the optimal choice.

2. Runtime and Memory Analysis

We provide a detailed inference runtime analysis of CarGait re-ranker across different K values in Fig. 3. Our method involves a certain level of complexity. However, in practice, the inference overhead is only ∼6.5 [msec] on a single A100 GPU with $K = 10$[1]. As mentioned in the paper, the re-ranker size is influenced by the feature map dimensions obtained by the single-stage model. Generally, the number of trainable parameters varies from 2.07M to 9.21M. At inference, the re-ranker has 0.4M parameters at most (compared to the size of single-stage gait models that varies from 2.5M to 13M). Training the model on four A100 GPUs with 40 [GB] of RAM takes approximately 16 hours.

3. Experiments on GREW

In the paper (Evaluation section), we demonstrate CarGait's superiority over the existing re-ranking methods [8, 10, 12] on the *Gait3D* [11] and *OU-MVLP* [9] datasets. Here, we provide an additional comparison on the *GREW* [13] dataset. As shown in Tab. 1, CarGait surpasses existing re-rankers across all five `methods` in both Rank-1 and Rank-5 accuracy.

4. Additional Experiments

In Table 1 of the paper, we exclude results for settings where checkpoints are unavailable. To enrich our evaluation, we independently trained the `SwinGait-3D` model on two datasets, using the official OpenGait implementation [4]. Although in *GREW* we were not able to reproduce the exact paper results, we provide CarGait performance gains on

[1] For comparison, the first-stage global retrieval takes 0.1 [msec] per probe.

Method	Publication	R1				R5			
		KR	LBR	GCR	CG	KR	LBR	GCR	CG
`GaitPart [2]`	CVPR 20	44.2	41.2	48.6	**52.5**	60.8	65.5	65.3	**67.5**
`GaitSet [1]`	AAAI 19	44.7	41.2	48.6	**52.0**	62.6	65.5	65.3	**68.0**
`GaitBase [5]`	CVPR 23	57.1	51.5	60.4	**67.2**	72.2	74.9	74.8	**78.5**
`DGV2-P3D [3]`	ArXiv 23	74.6	64.3	77.4	**79.2**	86.2	86.5	87.6	**88.7**
`SG++ [6]`	AAAI 24	84.2	80.4	86.0	**88.2**	93.0	93.4	93.0	**94.6**

Table 1. Rank-K accuracy [%] on *GREW* dataset [13] for different re-ranking methods: k-reciprocal (KR) [12], LBR [8], and GCR [10], compared to CarGait (CG). Best results are in bold.

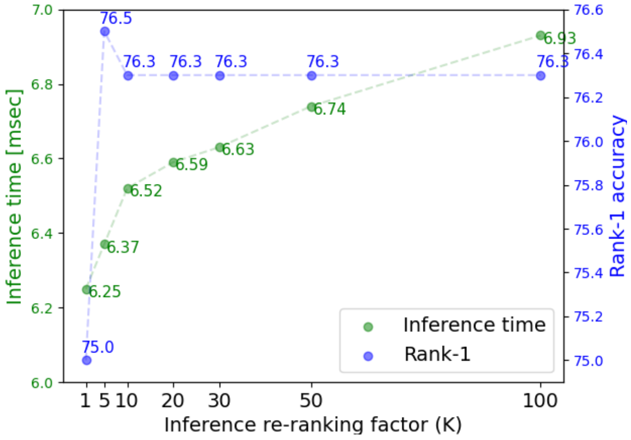

Figure 3. CarGait runtime analysis per probe with varying values of K. The results were obtained on a single A100 GPU using the `SwinGait-3D` model and the *Gait3D* dataset. The inference time per probe [in msec] is shown in green (left-hand side), while Rank-1 (R1) performance is depicted in blue (right-hand side).

both datasets (see Tab. 2).

Method	*GREW*				*OU-MVLP*	
	R1		R5		R1	
	Initial	CG	Initial	CG	Initial	CG
`SwinGait3D [3]`	78.7	**79.5**	88.6	**89.2**	91.1	**91.3**

Table 2. Rank-K accuracy [%] on additional settings.

5. Verification

In our evaluation, we adopt Rank-K and mAP, which are the standard metrics in gait recognition. Here, we highlight the relevance of an important complementary metric: **verification** performance, measured by TPR@FPR. This metric evaluates how reliably and securely a system can verify identities under strict error restrictions, which may be a critical requirement in real-world scenarios.

To this end, we use the *Gait3D* dataset (which includes multiple positives per probe), and the top $K = 1000$ single-stage predictions, to calculate TPR@FPR=1e-2 - before and after re-ranking. The results in Tab. 3 demonstrate consistent gains achieved by CarGait.

Method	Initial	CG
`GaitPart [2]`	21.0	**21.9**
`GaitGL [7]`	21.9	**27.5**
`GaitSet [1]`	27.1	**39.3**
`GaitBase [5]`	52.9	**59.5**
`DGV2-P3D [3]`	65.3	**67.4**
`SwinGait3D [3]`	67.9	**69.3**
`SG++ [6]`	69.7	**73.7**

Table 3. CarGait enhancements in verification performance (TPR@FPR=1e-2) on the *Gait3D* dataset [11] with $K = 1000$.

6. Ablations

Table 5 in the paper presents an ablation study using the `SwinGait-3D` model trained on the *Gait3D* dataset. Here, we extend this analysis with additional ablations focused on the cross-attention module. CarGait employs a single cross-attention block with 8 heads. As shown in Tab. 4, modifying the number of attention blocks or heads results in only a minor effect on performance.

Method	#Heads	#Blocks	R1	R5	mAP
Initial	–	–	75.0	86.7	66.69
H = 4	4	1	76.6	88.8	67.61
H = 16	16	1	76.2	89.2	67.84
B = 2	8	2	76.1	89.2	67.77
B = 4	8	4	76.3	88.8	67.65
CarGait	8	1	76.3	88.6	67.59

Table 4. Cross-attention ablations with `SwinGait-3D` model and *Gait3D* dataset. Rank-1 (R1), Rank-5 (R5), and mAP are reported.

References

[1] Hanqing Chao, Yiwei He, Junping Zhang, and Jianfeng Feng. Gaitset: Regarding gait as a set for cross-view gait

recognition. In *Proceedings of the AAAI conference on artificial intelligence*, pages 8126–8133, 2019. 2

[2] Chao Fan, Yunjie Peng, Chunshui Cao, Xu Liu, Saihui Hou, Jiannan Chi, Yongzhen Huang, Qing Li, and Zhiqiang He. Gaitpart: Temporal part-based model for gait recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14225–14233, 2020. 2

[3] Chao Fan, Saihui Hou, Yongzhen Huang, and Shiqi Yu. Exploring deep models for practical gait recognition, 2023. 1, 2

[4] Chao Fan, Junhao Liang, Chuanfu Shen, Saihui Hou, Yongzhen Huang, and Shiqi Yu. Opengait: Revisiting gait recognition towards better practicality. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9707–9716, 2023. 1

[5] Chao Fan, Junhao Liang, Chuanfu Shen, Saihui Hou, Yongzhen Huang, and Shiqi Yu. Opengait: Revisiting gait recognition towards better practicality. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9707–9716, 2023. 2

[6] Chao Fan, Jingzhe Ma, Dongyang Jin, Chuanfu Shen, and Shiqi Yu. Skeletongait: Gait recognition using skeleton maps. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 1662–1669, 2024. 2

[7] Beibei Lin, Shunli Zhang, and Xin Yu. Gait recognition via effective global-local feature representation and local temporal aggregation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14648–14656, 2021. 2

[8] Chuanchen Luo, Yuntao Chen, Naiyan Wang, and Zhaoxiang Zhang. Spectral feature transformation for person re-identification. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4976–4985, 2019. 1, 2

[9] Noriko Takemura, Yasushi Makihara, Daigo Muramatsu, Tomio Echigo, and Yasushi Yagi. Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition. *IPSJ transactions on Computer Vision and Applications*, 10:1–14, 2018. 1

[10] Yuqi Zhang, Qi Qian, Hongsong Wang, Chong Liu, Weihua Chen, and Fan Wang. Graph convolution based efficient reranking for visual retrieval. *IEEE Transactions on Multimedia*, 26:1089–1101, 2023. 1, 2

[11] Jinkai Zheng, Xinchen Liu, Wu Liu, Lingxiao He, Chenggang Yan, and Tao Mei. Gait recognition in the wild with dense 3d representations and a benchmark. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 20228–20237, 2022. 1, 2

[12] Zhun Zhong, Liang Zheng, Donglin Cao, and Shaozi Li. Re-ranking person re-identification with k-reciprocal encoding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1318–1327, 2017. 1, 2

[13] Zheng Zhu, Xianda Guo, Tian Yang, Junjie Huang, Jiankang Deng, Guan Huang, Dalong Du, Jiwen Lu, and Jie Zhou. Gait recognition in the wild: A benchmark. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 14789–14799, 2021. 1, 2