# GaitW: Enhancing Gait Recognition in the Wild using Dynamic Information

Daksh Thapar*[1], Jayesh Chaudhari*[2], Sunny Manchanda[3], Aditya Nigam[1], and Chetan Arora[2]

[1] Indian Institute of Technology Mandi, India
[2] Indian Institute of Technology Delhi, India
[3] Defense young Scientist Laboratory, India
d18033@students.iitmandi.ac.in, jayeshc.cstaff@iitd.ac.in
faculty.iitmandi.ac.in/~aditya/, www.cse.iitd.ac.in/~chetan/
sunny.dysl-ai@gov.in

# Supplementary Material

## A Subsampled GEI ($\mathcal{G}_s$)

Gait recognition faces significant challenges due to speed variations among individuals, a concern that complicates the analysis beyond intra-individual discrepancies to inter-individual variations observed across diverse video instances. To enhance the model's resilience to these speed variations, it's crucial to develop an ability to identify gait characteristics across various speeds. One strategy involves creating a representation of the gait at a faster speed and integrating the learning of its features with those of the standard speed gait sample. By selecting frames at a particular rate from a video of gait samples, we can generate an accelerated gait sequence (for example, by choosing every other frame). However, we opted not to directly incorporate this accelerated sequence (the selected frames) into our attention mechanism due to the high spatial and temporal complexity involved in processing attention. Moreover, adding the accelerated sequence to the attention module would lead to redundancy, as these frames are essentially included within the Frame Wise Silhouettes ($\mathcal{F}$).

Hence, we have generated subsampled gait energy image (`GEI`) from these sampled frames by averaging frames present in the accelerated gait sequence (refer to Fig. 1) named as $\mathcal{G}_s$. Given $\mathcal{F} = \{f_i \mid i = 1, 2, \ldots, T\}$, and $f_i \in \mathbb{R}^{H \times W}$ is the $i^{th}$ frame of the input gait sequence. We derive $\mathcal{G}_s = \frac{1}{T} \sum_{\substack{i=1 \\ i+=2}}^{T} f_i$. This strategy is particularly effective as $\mathcal{G}_s$ significantly diverges from the conventional Gait Energy Image (GEI), capturing unique speed-related information nuances. The rich, speed-specific insights within $\mathcal{G}_s$ offer a substantial boost to our model's robustness and its capacity to differentiate between gaits. Incorporating $\mathcal{G}_s$ into our model (by adding just one extra frame) is efficiently compatible with the attention mechanism. This addition contributes to a manageable increase in processing while significantly improving the model's consistency and accuracy in recognizing gaits across various speeds, as illustrated in Fig. 1.

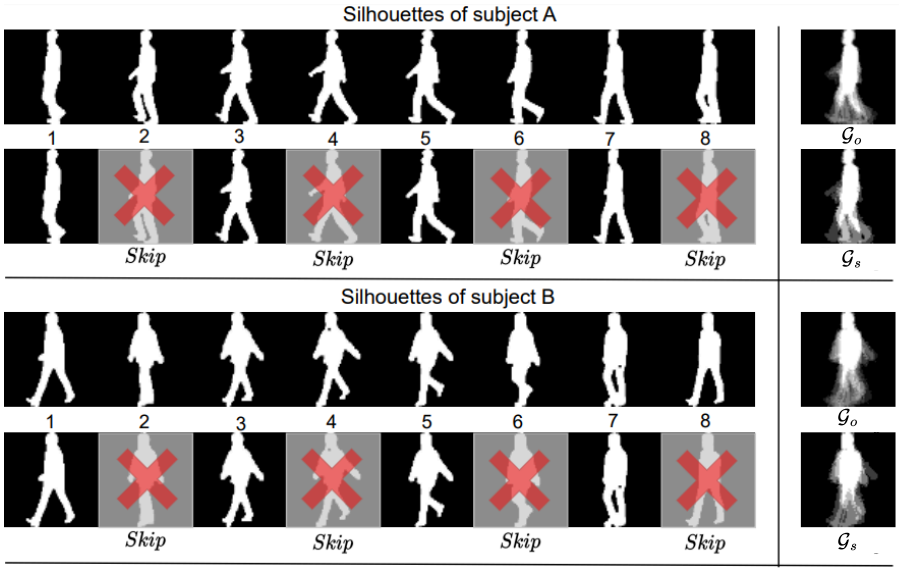


**Fig. 1: Subsample GEI:** First row shows that all frames selected for generating $\mathcal{G}_o$ and the respective generated $\mathcal{G}_o$. While the second row shows that the alternate frames are skipped for generating $\mathcal{G}_s$ and respective generated $\mathcal{G}_s$. One can observe that $\mathcal{G}_s$ captures the dynamic information representing the gait sample at accelerated speed so as to address speed variations.

## B GEI based Silhouette Mask Annealing

In gait recognition, not all regions within a silhouette are equally significant. It is the dynamic regions that harbour the essential characteristics for recognizing gait patterns. The `GEI` is computed as the average of all the silhouettes from a gait video, encapsulating holistic and dynamic information of the gait sequence. Hence, we exploit the dynamic cues present in the `GEI` to discern the dynamic (salient) regions within the silhouettes.

Each silhouette from the gait video represents a binary image, where pixel intensity is either 0 or 255. Therefore, when averaged across the entire video, the pixel values in the `GEI` can range between 0 and 255. If at a specific pixel, all silhouettes possessed a value of 255, the corresponding pixel in the `GEI` would also exhibit a value of 255, indicating that the pixel was static across the silhouettes. The region having such pixels is termed as foreground. In contrast, if at a certain pixel, some silhouettes had a value of 0 while others had a value of 255, the `GEI` pixel value would fall between 1 and 254, signifying a dynamic pixel due to the movement observed across the silhouettes. The region having such pixels is termed as boundary.

We derive an initial binary mask ($m_0$-`GEI`) from the `GEI` by setting the foreground regions to 0 and the boundary regions to 1. The background regions are already 0. This mask is then iteratively refined to $m_1$-`GEI` and $m_2$-`GEI` through the application of dilation (c.f. Fig. 2). Following the completion of Stage I and

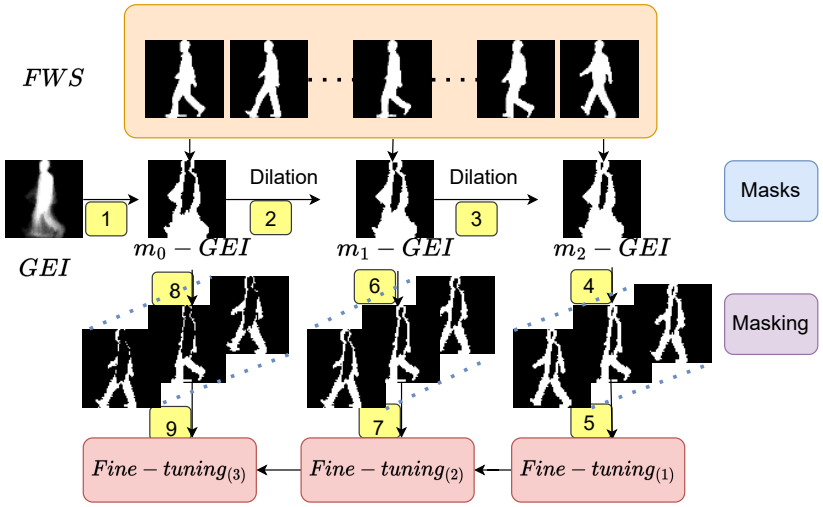


**Fig. 2:** The masking strategy utilised to focus on key regions. Firstly the `GEI` is computed from `FWSs`. Then we compute mask $m_0$ from `GEI` by setting static regions to 0 and dynamic regions to 1. This mask is refined to $m_1$ and $m_2$ by dilating $m_0$. Finally, each of these masks are applied to the `FWSs` and fine tuning happens in a step-wise manner, as marked under arrows in yellow color.

II training (refer section 4.1 and 4.2 of the paper), we further accentuate the model's focus on dynamic contours by generating silhouettes ($\mathcal{F}$) with the static regions' pixel intensity set to 0. This is achieved by multiplying the created masks ($m_0, m_1, m_2$-`GEI`) on each silhouette in a gait video to produce a masked silhouette.

The `GaitW` undergoes fine-tuning in a step-wise manner. Initially, the $m_2$-`GEI` mask is applied to the frames, and the model is fine-tuned using these masked frames. This fine-tuning process is subsequently repeated with the $m_1$-`GEI`, and finally, the ($m_0$-`GEI`) mask. Each step involves applying the respective mask to the silhouette frames, thereby progressively refining the model's focus on the most informative and dynamic regions of the gait. This iterative method ensures that `GaitW` increasingly emphasizes the critical dynamics of gait, such as limb movements while diminishing the impact of static elements.

| **Gallery NM#1-4** | **Angles from 0°-108°** | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Method** | 0° | 18° | 36° | 54° | 72° | 90° | 108° | 126° | 144° | 162° | 180° | Mean |
| Stage I and stage II only | 93.5 | 96.9 | 96.2 | 96.0 | 92.5 | 91.4 | 93.2 | 96.1 | 96.2 | 95.0 | 87.4 | 94.0 |
| $Fine - tuning_{(1)}$ ($m_2$ mask) | 94.0 | 97.1 | 96.5 | 96.4 | 93.2 | 92.1 | 93.5 | 96.5 | 96.4 | 95.0 | 88.3 | 94.4 |
| $Fine - tuning_{(2)}$ ($m_1$ mask) | 94.0 | 97.3 | 96.8 | 96.3 | 93.5 | 92.4 | 93.6 | 96.4 | 96.4 | 95.1 | 88.5 | 94.6 |
| $Fine - tuning_{(3)}$ ($m_0$ mask) | 94.2 | 97.7 | 97.0 | 96.3 | 93.7 | 92.1 | 93.8 | 96.7 | 96.8 | 95.4 | 89.7 | 94.9 |

**Table 1:** Rank-1 accuracy (%) on `CASIA-B` on all angles, CL#1-2 conditions, under LT-74 setting of three step-wise mask fine-tuning steps and the only stage I and stage II training.

With each step of fine-tuning, there is an observed enhancement in the accuracy of the GaitW, as documented in Tab. 1, highlighting the improvement in accuracy following each fine-tuning phase.

## C Complete results on CASIA-B

**Data-set Specifications:** Due to space constrains we have excluded some results (ST and MT settings) on CASIA-B, which we are reporting in Tab. 3, Tab. 5 and Tab. 4. The CASIA-B dataset has total 124 subjects and for each subject data is collected in 10 groups. Six groups are classified to normal walking condition (NM01-NM06). Two of the groups are classified as person wearing coat as a walking condition (CL01-CL02) and remaining two are are classified into person holding bag as a walking condition (BG01-BG02). Each group is further divided into 11 angle, for cross view angle setting ranging from 0° to 180° with an interval of 18°. Hence, there are a total of 124 subjects × 10 groups × 11 view angles = 13,640 gait sequences in CASIA-B.
**Training and testing methodology:** Officially there is no specific training-test split defined for CASIA-B. In order to have fair comparisons, we have conducted experiments on three settings as suggested and utilized in most SOTA techniques [1,6]. We name them as Small-sample Training (ST), Medium-sample Training (MT) and Large-sample Training (LT). In ST (referred to as ST24), the first 24 subjects (labelled from 001-024) are used for training, and the remaining 100 subjects are left for testing. In MT (referred to as MT62), the first 62 subjects are used for training, and the rest 62 subjects are left for testing. In LT (referred to as LT74), the first 74 subjects are used for training, and the remaining 50 subjects are left for testing. In the testing stage, sequences NM01-NM04 constitute the gallery set, while sequences NM05-NM06, BG01-BG02, and CL01-CL02 form the probe set, facilitating the evaluation of performance.
**Comparative analysis:** The GaitW has outperformed in the most popular setting LT74 for CL (included in the main paper) and comparable results for NM and BG as demonstrated in Tab. 5. For NM condition GaitW has performance of 98.8% and current SOTA (MSGR [5]) has perfromance of 99.1% and for BG condition GaitW has performance of 97.1 and the current SOTA (MSGR [5]) has performance of 97.6%. It is interesting to note that MSGR uses silhouettes and pose modality jointly to achieve this results. GaitW beats current silhouettes SOTA (HSTGait [3]) in NM condition by 0.4% and beats current silhouettes SOTA (DyGait [4]) in BG condition by 0.9%. In other two setting MT62 and ST24, GaitW outperformed the current SOTA GaitGL [2]. For MT62, across all conditions (NM, BG and CL) GaitW achieves 5.54% better accuracy than GaitGL(refer Tab. 4). Similarly for ST24, across all conditions (NM, BG and CL) GaitW achieves 15.3% better accuracy than GaitGL (refer Tab. 3).

## D Parameters Comparison

We have compared the number of parameters (#Params), FLOPs, gait signature length (Sig. len.), along with accuracy (Acc.) for our proposed model alongside state-of-the-art (SOTA) models on CASIA-B dataset is given in the table below.

| Method | #Params | FLOPs | Sig. len. | Acc. |
|---|---|---|---|---|
| GaitGL | 11.19 | 58.55 | 1024 | 93.5 |
| GaitBase | 7.30 | 9.45 | 256 | 89.6 |
| CSTL | 9.09 | 26.2 | - | 93.4 |
| LangGait | - | 66.2 | 512 | 92.3 |
| HSTGait | - | 38.4 | - | 94.3 |
| GaitW | 15.2 | 39.6 | 128 | 94.9 |

**Table 2:** Parametric capacity comparison across different methods on OU-MVLP dataset.

GaitW is engineered for balanced accuracy and robustness across various datasets. Existing literature often lacks consistent reporting of parameters and FLOPs, and the unavailability of the code for some SOTA models restricts comparison. (refer Tab. 2)

## E Testing protocol for GREW-1K

In GREW test dataset, each subject has two labeled samples, while all the probe samples are unlabeled. The evaluation has to be done online through their server. Hence, we curated GREW-1K dataset for easier ablation study. We took first 1,000 subjects from GREW and used only their two labeled samples. We used the first sample as the gallery and the second as the probe, and then performed 1:N matching to measure performance.

## F Qualitative Analysis

In order to justify that our hardness module works consistently, we have chosen 10 subjects randomly. We have used pre-trained GaitW, GaitSet and SwinV2 models and extract the embeddings for 4 easy and 4 hard samples per subject. A t-sne [2] plot for such 40 easy and hard samples are row-wise, shown in Fig. 3. The row 1, shows the plot of model GaitW, GaitSet, and SwinV2 on easy samples designated as per our curriculum. The row 2, shows the plot of the same three models over hard samples as per our curriculum. It is evident that over easy samples, all model works well. One can observe that GaitW is able to differentiate between subjects better, especially when the samples are hard, as compared to the other models.

| Gallery NM#1-4 | | | Angles from 0°-108° | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Probe | | | 0° | 18° | 36° | 54° | 72° | 90° | 108° | 126° | 144° | 162° | 180° | Mean |
| NM#5-6 | ViDP | Jour IEEE'13 | - | - | - | 59.1 | - | 50.2 | - | 57.5 | - | - | - | - |
| | CMCC | Jour IEEE'13 | 46.3 | - | - | 52.4 | - | 48.3 | - | 56.9 | - | - | - | - |
| | CNN-LB | Jour IEEE'16 | 54.8 | - | - | 77.8 | - | 64.9 | - | 76.1 | - | - | - | - |
| | GaitSet | AAAI'19 | 64.6 | 83.3 | 90.4 | 86.5 | 80.2 | 75.5 | 80.3 | 86.0 | 87.1 | 81.4 | 59.6 | 79.5 |
| | GaitGL | ICCV'21 | 77.0 | 87.8 | 93.9 | 92.7 | 83.9 | 78.7 | 84.7 | 91.5 | 92.5 | 89.3 | 74.4 | 86.0 |
| | GaitW | | **91.4** | **95.6** | **97.8** | **95.2** | **92.6** | **93.2** | **92.7** | **93.6** | **95.8** | **95.7** | **89.4** | **93.9** |
| BG#1-2 | GaitSet | AAAI'19 | 55.8 | 70.5 | 76.9 | 75.5 | 69.7 | 63.4 | 68.0 | 75.8 | 76.2 | 70.7 | 52.5 | 68.6 |
| | GaitGL | ICCV'21 | 68.1 | 81.2 | 87.7 | 84.9 | 76.3 | 70.5 | 76.1 | 84.5 | 87.0 | 83.6 | 65.0 | 78.6 |
| | GaitW | | **84.6** | **90.1** | **92.3** | **93.2** | **89.6** | **87.7** | **85.8** | **92.3** | **94.8** | **91.5** | **85.0** | **89.7** |
| CL#1-2 | GaitSet | AAAI'19 | 29.4 | 43.1 | 49.5 | 48.7 | 42.3 | 40.3 | 44.9 | 47.4 | 43.0 | 35.7 | 25.6 | 40.9 |
| | GaitGL | ICCV'21 | 46.9 | 58.7 | 66.6 | 65.4 | 58.3 | 54.1 | 59.5 | 62.7 | 61.3 | 57.1 | 40.6 | 57.4 |
| | GaitW | | **81.6** | **87.8** | **88.4** | **86.2** | **82.5** | **84.3** | **83.2** | **83.0** | **87.6** | **84.3** | **79.8** | **84.4** |

**Table 3:** Rank-1 accuracy (%) on CASIA-B under all view angles, different conditions, with ST-24 setting. Refer to the main text for details on testing protocol.

| Gallery NM#1-4 | | | Angles from 0°-108° | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Probe | | | 0° | 18° | 36° | 54° | 72° | 90° | 108° | 126° | 144° | 162° | 180° | Mean |
| NM#5-6 | AE | Jour Neuro.'17 | 49.3 | 61.5 | 64.4 | 63.6 | 63.7 | 58.1 | 59.9 | 66.5 | 64.8 | 56.9 | 44.0 | 59.3 |
| | MGAN | Jour IEEE'19 | 54.9 | 65.9 | 72.1 | 74.8 | 71.1 | 65.7 | 70.0 | 75.6 | 76.2 | 68.6 | 53.8 | 68.1 |
| | GaitSet | AAAI'19 | 86.8 | 95.2 | 98.0 | 94.5 | 91.5 | 89.1 | 91.1 | 95.0 | 97.4 | 93.7 | 80.2 | 92.0 |
| | GaitGL | ICCV'21 | 93.9 | 97.6 | 98.8 | 97.3 | 95.2 | 92.7 | 95.6 | 98.1 | **98.5** | 96.5 | 91.2 | 95.9 |
| | GaitW | | **95.1** | **98.7** | **99.4** | **98.0** | **97.3** | **95.1** | **97.0** | **98.6** | 98.3 | **97.1** | **93.4** | **97.1** |
| BG#1-2 | AE | Jour Neuro.'17 | 29.8 | 37.7 | 39.2 | 40.5 | 43.8 | 37.5 | 43.0 | 42.7 | 36.3 | 30.6 | 28.5 | 37.2 |
| | MGAN | Jour IEEE'19 | 48.5 | 58.5 | 59.7 | 58.0 | 53.7 | 49.8 | 54.0 | 51.3 | 59.5 | 55.9 | 43.1 | 54.7 |
| | GaitSet | AAAI'19 | 79.9 | 89.9 | 91.2 | 86.7 | 81.6 | 76.7 | 81.0 | 88.2 | 90.3 | 88.5 | 73.0 | 84.3 |
| | GaitGL | ICCV'21 | 88.5 | 95.1 | 95.9 | 94.2 | 91.5 | 85.4 | 89.0 | 95.4 | 97.4 | 94.3 | 86.3 | 92.1 |
| | GaitW | | **90.4** | **96.9** | **97.3** | **95.2** | **94.7** | **91.3** | **92.8** | **96.1** | **97.6** | **94.8** | **87.0** | **94.0** |
| CL#1-2 | AE | Jour Neuro.'17 | 18.7 | 21.0 | 25.0 | 25.1 | 25.0 | 26.3 | 28.7 | 30.0 | 23.6 | 23.4 | 19.0 | 24.2 |
| | MGAN | Jour IEEE'19 | 23.1 | 34.5 | 36.3 | 33.3 | 32.9 | 32.7 | 34.2 | 37.6 | 33.7 | 26.7 | 21.0 | 31.5 |
| | GaitSet | AAAI'19 | 52.0 | 66.0 | 72.8 | 69.3 | 63.1 | 61.2 | 63.5 | 66.5 | 57.5 | 60.0 | 45.9 | 62.5 |
| | GaitGL | ICCV'21 | 70.7 | 83.2 | 87.1 | 84.7 | 78.2 | 71.3 | 78.0 | 83.7 | 83.6 | 77.1 | 63.1 | 78.3 |
| | GaitW | | **92.3** | **95.6** | **95.7** | **94.8** | **93.2** | **89.1** | **90.0** | **92.3** | **91.7** | **90.4** | **85.3** | **91.8** |

**Table 4:** Rank-1 accuracy (%) on CASIA-B under all view angles, different conditions, with MT-62 setting. Refer to the main text for details on testing protocol.

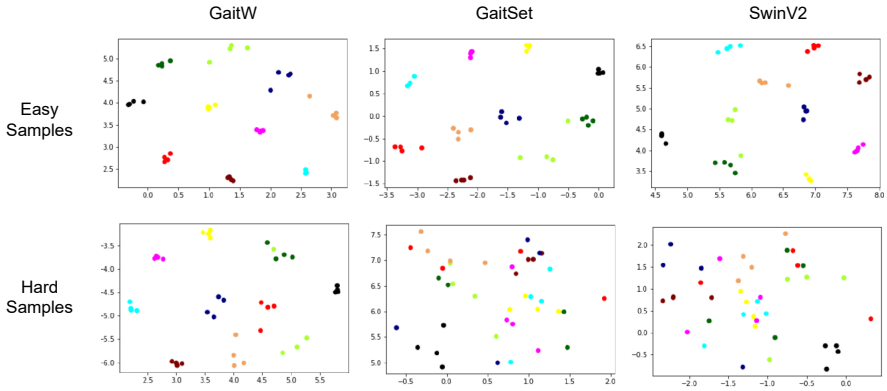


**Fig. 3:** Row 1, shows $t - sne$ plots of three models: `GaitW`, GaitSet, and SwinV2, for 40 samples (taken from 10 subjects) declared easy by our hardness scoring module. Similarly, in Row 2, 40 hard samples are plotted. One can see that over easy samples all models can discriminate but on hard samples `GaitW` provides best discriminative features.

| Gallery NM#1-4 | | | Angles from 0°-108° | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Probe | | | 0° | 18° | 36° | 54° | 72° | 90° | 108° | 126° | 144° | 162° | 180° | Mean |
| NM#5-6 | GaitPart | CVPR'20 | 94.1 | 98.6 | 99.3 | 98.5 | 94.0 | 92.3 | 95.9 | 98.4 | 99.2 | 97.8 | 90.4 | 96.2 |
| | GLN | ECCV'20 | 93.20 | 99.30 | 99.50 | 98.70 | 96.10 | 95.60 | 97.20 | 98.10 | 99.30 | 98.60 | 90.10 | 96.88 |
| | 3DLocal | ICCV'21 | 96.0 | 99.0 | 99.5 | 98.9 | 97.1 | 94.2 | 96.3 | 99.0 | 98.8 | 98.5 | 95.2 | 97.5 |
| | CSTL | ICCV'21 | 97.8 | 99.4 | 99.2 | 98.4 | 97.3 | 95.2 | 96.7 | 98.9 | 99.4 | 99.3 | 96.7 | 98.0 |
| | SRN+CB | TBBIS'21 | 94.4 | 99.3 | 99.4 | 98.7 | 96.8 | 96.8 | 97.5 | 98.5 | 99.5 | 98.8 | 92.3 | 97.5 |
| | GaitGL | ArXiv'22 | 96.6 | 98.8 | 99.1 | 98.1 | 97.0 | 96.8 | 97.9 | 99.2 | 99.3 | 99.3 | 95.6 | 98.0 |
| | LangGait | CVPR'22 | 95.7 | 98.1 | 99.1 | 98.3 | 96.4 | 95.2 | 97.5 | 99.0 | 99.3 | 98.9 | 94.9 | 97.5 |
| | MetaGait | ECCV'22 | t97.3 | 99.2 | 99.5 | 99.1 | 97.2 | 95.5 | 97.6 | 99.1 | 99.3 | 99.1 | 96.7 | 98.1 |
| | GaitGCI-L | CVPR'23 | - | - | - | - | - | - | - | - | - | - | - | 98.4 |
| | DANet | CVPR'23 | 96.4 | 99.1 | 99.2 | 98.2 | 96.6 | 95.5 | 97.6 | 99.4 | 99.5 | 99.3 | 96.9 | 98.0 |
| | MMGaitf. | CVPR'23 | 98.1 | 98.6 | 99.0 | 98.1 | 98.4 | 97.8 | 98.1 | 99.0 | 99.2 | 99.1 | 97.3 | 98.4 |
| | GaitBase | CVPR'23 | - | - | - | - | - | - | - | - | - | - | - | 97.6 |
| | GaitRef | IJCB'23 | 97.2 | 98.7 | 99.1 | 98.0 | 97.3 | 97.0 | 98.0 | 99.4 | 99.4 | 98.9 | 96.4 | 98.1 |
| | STANet | ICCV'23 | 96.4 | 99.4 | 99.3 | 98.9 | 97.0 | 95.8 | 98.2 | 99.2 | 99.6 | 99.2 | 96.0 | 98.1 |
| | DyGait | ICCV'23 | 97.4 | 98.9 | 99.2 | 98.3 | 97.7 | 96.8 | 98.2 | 99.3 | 99.3 | 99.2 | 97.6 | 98.4 |
| | HSTGait | ICCV'23 | 97.6 | 98.0 | 99.6 | 98.2 | 97.4 | 96.5 | 97.9 | 99.3 | 99.4 | 98.4 | 97.0 | 98.1 |
| | MSGR | TMM'23 | **99.3** | 99.2 | 99.2 | 99.1 | **99.0** | **99.0** | **99.5** | **99.7** | **99.5** | **99.5** | **98.3** | **99.2** |
| | MSAFF | IJCB'23 | 99.1 | **99.4** | 99.3 | 99.1 | 98.9 | 98.9 | 98.9 | 99.2 | 99.7 | 99.6 | 97.8 | 99.1 |
| | QAGait | AAAI'24 | - | - | - | - | - | - | - | - | - | - | - | 97.9 |
| | CLASH | TIP'24 | - | - | - | - | - | - | - | - | - | - | - | 98.3 |
| | GaitW | | 98.3 | 98.9 | **99.8** | **99.1** | 98.7 | 98.0 | 98.2 | 99.4 | 99.3 | 99.4 | 97.6 | 98.8 |
| BG#1-2 | GaitPart | CVPR'20 | 89.1 | 94.8 | 96.7 | 95.1 | 88.3 | 94.9 | 89.0 | 93.5 | 96.1 | 93.8 | 85.8 | 91.5 |
| | GLN | ECCV'20 | 91.10 | 97.68 | 97.78 | 95.20 | 92.50 | 91.20 | 92.40 | 96.00 | 97.50 | 94.95 | 88.10 | 94.04 |
| | 3DLocal | ICCV'21 | 94.7 | 98.7 | 98.8 | 97.5 | 93.3 | 91.7 | 92.8 | 96.5 | 98.1 | 97.3 | 90.7 | 95.5 |
| | CSTL | ICCV'21 | 95.0 | 96.8 | 97.9 | 96.0 | 94.0 | 90.5 | 92.5 | 96.8 | 97.9 | 99.0 | 94.3 | 95.4 |
| | SRN+CB | TBBIS'21 | 91.5 | 97.4 | 98.4 | 97.1 | 92.2 | 89.7 | 93.1 | 96.2 | 97.5 | 96.5 | 88.0 | 94.3 |
| | GaitGL | ArXiv'22 | 93.9 | 97.3 | 97.6 | 96.2 | 94.7 | 91.0 | 94.4 | 97.2 | 98.6 | 97.1 | 91.6 | 95.4 |
| | LangGait | CVPR'22 | 94.2 | 96.2 | 96.8 | 95.8 | 94.3 | 89.5 | 91.7 | 96.8 | 98.0 | 97.0 | 90.9 | 94.6 |
| | MetaGait | ECCV'22 | 92.9 | 96.7 | 97.1 | 96.4 | 94.7 | 90.4 | 92.9 | 97.2 | 98.5 | 98.1 | 92.3 | 95.2 |
| | GaitGCI | CVPR'23 | - | - | - | - | - | - | - | - | - | - | - | 96.6 |
| | MMGaitf. | CVPR'23 | 97.1 | 95.9 | 97.1 | 95.7 | 96.1 | 95.2 | 95.2 | 97.1 | 97.3 | 96.1 | 93.5 | 96.0 |
| | DANet | CVPR'23 | 95.0 | 97.3 | 98.3 | 97.4 | 94.7 | 91.0 | 93.9 | 97.4 | 98.2 | 97.6 | 94.2 | 95.9 |
| | GaitBase | CVPR'23 | - | - | - | - | - | - | - | - | - | - | - | 94.0 |
| | GaitRef | IJCB'23 | 94.4 | 96.4 | 97.3 | 96.8 | 96.2 | 92.2 | 94.4 | 97.2 | 98.7 | 97.9 | 93.3 | 95.9 |
| | STANet | ICCV'23 | 94.4 | 98.2 | 98.9 | 97.5 | 94.1 | 91.2 | 93.9 | 97.4 | 98.5 | 97.8 | 94.0 | 96.0 |
| | DyGait | ICCV'23 | 94.5 | 96.9 | 97.4 | 96.1 | 95.4 | 94.0 | 94.8 | 97.6 | 98.5 | 97.7 | 94.9 | 96.2 |
| | HSTGait | ICCV'23 | 95.0 | 96.5 | 97.3 | 96.6 | 95.3 | 93.3 | 94.6 | 96.8 | 98.6 | 97.7 | 92.9 | 95.9 |
| | MSGR | IJCB'23 | **98.3** | 97.9 | 98.1 | 97.4 | **96.9** | **95.6** | **97.3** | **98.5** | **99.1** | **98.3** | **96.4** | **97.6** |
| | MSAFF | TIP'24 | 97.7 | **98.5** | 98.6 | 98 | **96.9** | 95.3 | 96.2 | 97.6 | 98.5 | 97.7 | 94.1 | 97.1 |
| | QAGait | AAAI'24 | - | - | - | - | - | - | - | - | - | - | - | 94.6 |
| | CLASH | TIP'24 | - | - | - | - | - | - | - | - | - | - | - | 95.3 |
| | GaitW | | 97.5 | 98.3 | **99.0** | **97.6** | 96.3 | 95.0 | 95.5 | 97.6 | 98.9 | 97.3 | 94.4 | 97.1 |

**Table 5:** Rank-1 accuracy (%) on CASIA-B under all view angles, nm and bg conditions, with LT-74 setting. Refer to the main text for details on testing protocol.

## References

1. Lin, B., Zhang, S., Bao, F.: Gait recognition with multiple-temporal-scale 3d convolutional neural network. In: Proceedings of the 28th ACM international conference on multimedia. pp. 3054–3062 (2020) 4
2. Van der Maaten, L., Hinton, G.: Visualizing data using t-sne. Journal of machine learning research **9**(11) (2008) 4, 5
3. Wang, L., Liu, B., Liang, F., Wang, B.: Hierarchical spatio-temporal representation learning for gait recognition. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 19639–19649 (October 2023) 4
4. 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. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 13424–13433 (October 2023) 4
5. Wang, R., Shi, Y., Ling, H., Li, Z., Zhao, C., Wei, B., Li, H., Li, P.: Gait recognition with multi-level skeleton-guided refinement. IEEE Transactions on Multimedia pp. 1–12 (2023). https://doi.org/10.1109/TMM.2023.3323887 4
6. Wolf, T., Babaee, M., Rigoll, G.: Multi-view gait recognition using 3d convolutional neural networks. In: 2016 IEEE international conference on image processing (ICIP). pp. 4165–4169. IEEE (2016) 4