**Citation**: Guizhi Li, Weiwei Fang. A global-local part-shift network for gait recognition. *Journal of Harbin Institute of Technology (New Series)*. DOI:10.11916/j.issn.1005-9113.24064

# A Global-Local Part-Shift Network for Gait Recognition

*Guizhi Li* and Weiwei Fang*

(*School of Computer Science, Beijing Information Science and Technology University, Beijing 100192, China*)

**Abstract**: Gait recognition, a promising biometric technology, relies on analyzing individuals' walking patterns and offers a non-intrusive and convenient approach to identity verification. However, gait recognition accuracy is often compromised by external factors such as changes in viewpoint and attire, which present substantial challenges in practical applications. To enhance gait recognition performance under diverse viewpoints and complex conditions, a global-local part-shift network is proposed in this paper. This framework integrates two novel modules: the part-shift feature extractor and the dynamic feature aggregator. The part-shift feature extractor strategically shifts body parts to capture the intrinsic relationships between non-adjacent regions, enriching the recognition process with both global and local spatial features. The dynamic feature aggregator addresses long-range dependency issues by incorporating multi-range temporal modeling, effectively aggregating information across parts and time steps to achieve a more robust recognition outcome. Comprehensive experiments on the CASIA-B dataset demonstrate that the proposed global-local part-shift network delivers superior performance compared with state-of-the-art methods, highlighting its potential for practical deployment.

**Keywords**: gait recognition; global-local feature; part-shift

**CLC number**: TP394 **Document code**: A **Article ID**: 1005-9113(2025)00-0000-08

## 0 Introduction

As a technique used to recognize a person based on the distinctive way they walk, Gai recognition has gained increasing interest recently because of its possible uses across various domains such as security, forensics, and healthcare. From an individual's gait, it is possible to identify them from a distance and without requiring any physical contact.

While gait recognition has many potential benefits, it also poses some challenges. Gait patterns can be affected by camera view points and clothing, making it difficult to accurately identify individuals.

In order to solve these problems, many methods based on deep learning have been proposed. He et al.[1] proposed to use multi-task generative adversarial learning network to learn the feature representation of a specific perspective, and improve the representation ability of features. GaitSet[2] regard the gait sequence as an unordered set, and obtain more discrim-inative features through the designed global and local feature extraction dual-branch structure, and finally use pyramid pooling for feature mapping to get the final representation. In order to obtain more fine-grained spatial information, GaitPart[3] was proposed to divide the gait picture into different parts and perform feature extraction separately, and then model the temporal features through the proposed Micro-Motion Capture Module, and uses the attention mechanism to obtain short-term temporal features and eliminate redundant long-range information.

These methods only focus on the local information of the specific receptive field of the convolution kernel size on a single image, ignoring the internal feature relations between different parts. Therefore, to mine the internal relationship between different parts, improve the global perception ability while maintaining the local perception ability of the feature, a global-local part-shift network is proposed in this study to improve the discrimination ability of the feature:

1) We propose a part-shift module to recombine the features between different parts, and then perform feature extraction to improve the global perception ability of the module.

2) For the global-local features obtained in

Received 2024-10-27.

* Corresponding author: Guizhi Li, ph.D, Senior Experimentalist. Email.lgz403@sina.com.

different parts, we propose a dynamic temporal aggregator to model temporal features. This module can not only model short-range features, but also obtain long-range features. which can alleviate the long-range dependency problem, and improve the ability of temporal modeling.

3) Extensive experiments on the CASIA-B dataset demonstrate the superiority of our method.

## 1 Related Work

Current gait recognition methods based on deep learning can broadly be categorized into model-based methods[4-6] and appearance-based methods[2-3,7-8]. Below, we review and critically analyze representative methods from both categories, explicitly highlighting their limitations and clearly positioning our method in relation to them.

### 1.1 Model-Based Methods

Model-based approaches primarily rely on extracting structured features such as human pose or body joints to achieve gait recognition. For instance, PoseGait[5] leverages 3D human pose estimation to effectively mitigate the impact of viewpoint and clothing variations. It combines CNN and LSTM to jointly model spatial and temporal information and employs a multi-loss strategy for optimization. GaitGraph[6] utilizes 2D pose estimation and extracts spatial features via graph convolutional neural networks (GCNs), thus effectively capturing the structural dependencies among body joints. Furthermore, the approach proposed in Ref.[4] uses a multi-linear human model and fine-tunes a pre-trained human mesh recovery (HMR) network for gait recognition. Despite the effectiveness of these methods in handling pose variations, they are highly dependent on accurate pose estimation. Pose inaccuracies or occlusions can severely deteriorate their performance. Additionally, these methods usually focus primarily on structural information while neglecting richer appearance-based cues.

### 1.2 Appearance-Based Methods

Appearance-based methods directly extract features from silhouette images or gait sequences without explicitly modeling the human pose[9-16]. A representative work, Gait Energy Image (GEI)[17], compresses temporal gait information into a single template image, effectively representing individual walking patterns. However, the primary drawback of GEI-based methods, as pointed out in Ref.[18], is the substantial loss of temporal dynamics. GEnI[18] addresses robustness concerns by proposing gait entropy images, but temporal information loss persists. Extending GEI, Periodic Energy Image (PEI)[19] proposes a multi-channel gait template for capturing more discriminative features via adversarial training, yet this method remains view-dependent and still compromises temporal granularity[20-24]. To alleviate temporal information loss, recent methods utilize raw gait frame sequences as network inputs. The gait lateral network structure in Ref.[7] effectively captures discriminative compact features directly from gait contour sequences, significantly reducing feature redundancy without accuracy loss. Meanwhile, GaitPart[3] emphasizes local spatial-temporal characteristics by segmenting gait sequences into distinct body parts to extract fine-grained features. GaitSlice[8] further enhances subtle spatial interactions among adjacent gait parts through its SED module, incorporating a frame attention mechanism for temporal feature selection. Although these appearance-based methods have made considerable progress, a common limitation is their predominant reliance on local receptive fields via convolution operations, which restricts them to localized spatial-temporal interactions. Consequently, they fail to explicitly model the inherent global relationship among different gait parts.

Different from the aforementioned methods, our proposed global-local part-shift network explicitly addresses these limitations. Specifically, our part-shift module strategically recombines features across distinct gait parts, capturing intrinsic relationships between non-adjacent regions. This approach enhances global spatial perception while simultaneously preserving local feature integrity. To further mitigate limitations in temporal modeling, particularly concerning long-range dependencies, we introduce a dynamic feature aggregator. This module employs multi-range temporal modeling to effectively integrate both short-range and long-range temporal interactions within gait sequences, thus significantly improving the robustness and discriminative power of the recognition process.

## 2 Method

In this section, we first provide a brief introduction of our proposed approach, followed by a

detailed description of the Part-Shift Feature Extractor (PSFE) module, and finally present the Dynamic Feature Aggregator (DFA) module.

### 2.1 Overview

As shown in Fig. 1, our network receives a sequence of gait contours $x_i$ as input, where $i \in 1,2,\cdots,T$, $T$ denotes the number of frames. $x_i$ first is input to the PSFE, the PSFE enhances the richness of features by mining the internal relations of different body parts, and obtains the global part-aware vector $\boldsymbol{P}_g$ and local part-aware vector $\boldsymbol{P}_l$:

$$P_{gi}, P_{li} = \mathrm{PSFE}(x_i) \tag{1}$$

Afterwards, through the Horizontal Pooling (HP) operation (it is worth noting that the maximum function is used as the horizontal pooling strategy in this study), the features are mapped to obtain the decisive spatial feature vector $\boldsymbol{D}_g$ and $\boldsymbol{D}_l$:

$$D_{gi}, D_{li} = \mathrm{HP}(P_{gi}, P_{li}) \tag{2}$$

The decisive spatial features are sent to the DFA. The DFA contains two sub-modules, Multi-Range Modeling (MRM) and Feature Aggregator (FA). MRM maps features into short-range temporal features $T_s$ and long-range temporal features $T_l$ in the temporal dimension:

$$T_{s_{mi}}, T_{l_{mi}} = \mathrm{MRM}(D_{mi}) \tag{3}$$

where $m = g,l$, that is, the global salient spatial features $D_{gi}$ and local salient spatial features $D_{li}$ will be sent to the MRM for temporal modeling, but they do not share parameters.

Then, the global-local multi-range temporal information is sent to the FA for mapping to obtain the fusion feature $F_i$:

$$F_i = \mathrm{FA}(T_{m_n i}) \tag{4}$$

where $m = s,l; n = g,l$, the global-local long-range and short-range four types of features are used as input.

Finally, map $F_i$ through seperate Fully Connected Layer (FC) to get the final feature representation $X_i$:

$$X_i = \mathrm{FC}(F_i) \tag{5}$$

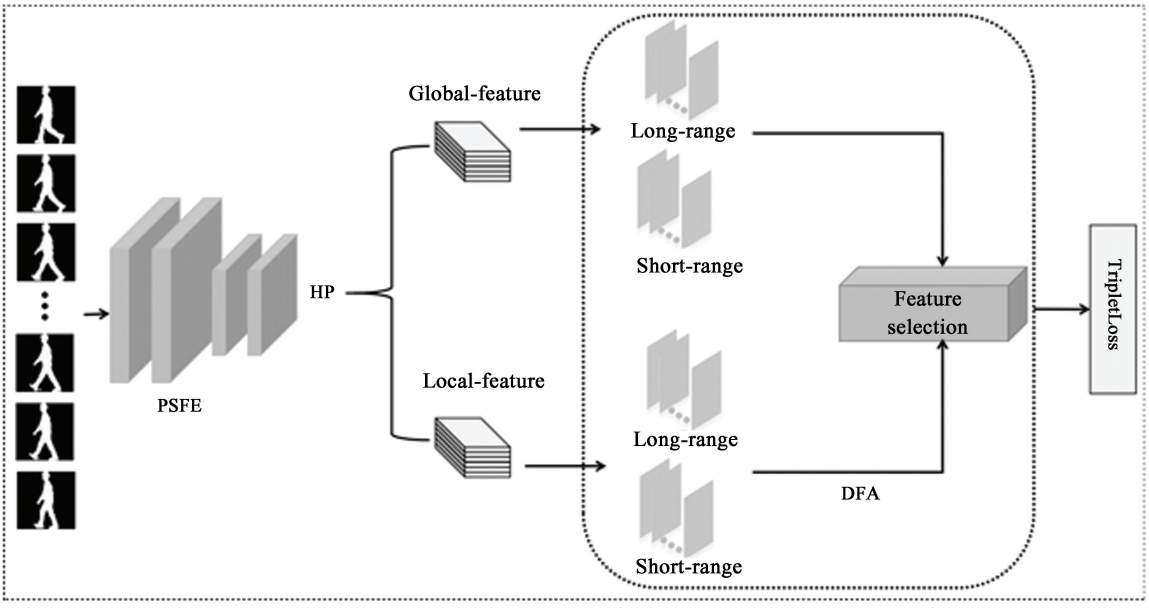



**Fig. 1 Overview of the proposed method**

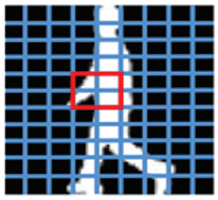
### 2.2 Part-Shift Feature Extractor (PSFE)

The part-shift Extractor aims to improve the perception of long-range spatial information while retaining the ability to extract fine-grained spatial features. It consists of two Part-Shift Convolution (PSConv) modules. The PSConv is introduced in detail below.

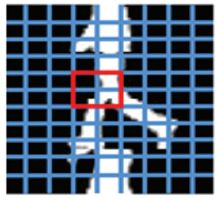
There are primarily two approaches to frame-level spatial feature extraction: ordinary 2D convolution method and part-based method. The part-based approach involves segmenting the input frame into distinct regions, followed by the application of 2D convolution within each region. However, both methods are limited to capturing local spatial features, lacking the capacity to model comprehensive spatial relationships across the entire frame. To address this

limitation, we propose an apart-shift convolution module grounded in an apart-aware approach, which facilitates the exchange of positions across segmented regions. As illustrated in Fig. 2, part-shift enables previously non-adjacent body parts to become adjacent, allowing convolutional operations to explicitly capture interconnections among these parts and enhance global spatial modeling.

The specific operation is illustrated in Fig. 3. The input feature map is divided into distinct segments, referred to as gait stripes. Initially, the sequentially arranged raw feature maps are processed by a convolutional neural network to model features within each segment. Subsequently, the part-shift operation reorders the feature maps across different segments.



(a) Original part feature



(b) Feature after part-shift transformation

**Fig. 2 After the part-shift change, the comparison of the change of the perception range**

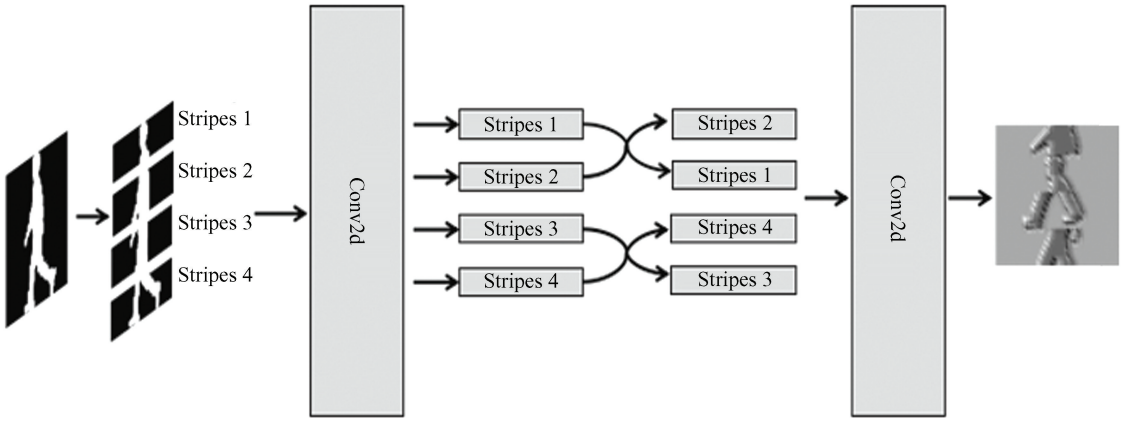



**Fig. 3 The detailed structure of PSConv module**

After exchanging the feature map, we model the original non-adjacent feature points through 2D convolution to obtain the intrinsic connection between long-distance spatial feature points.

A part-shift convolution module can be expressed as follows:

$$F_i = \mathrm{Conv2d}(\mathrm{PartShift}(\mathrm{Conv2d}(x_i))) \quad (6)$$

where $i \in 1,2,\cdots,T$, $T$ denotes the frame number of input gait sequence.

## 2.3 Dynamic Temporal Feature Aggregator

The dynamic feature aggregator further performs temporal modeling of global-local, frame-level spatial features across various granularities to address challenges related to long-range dependencies. This module comprises two independent multi-range feature aggregators, which do not share parameters, as well as the final feature aggregator. A detailed introduction to these modules will be provided subsequently.

Unlike other biometric recognition techniques, gait recognition uniquely requires feature modeling across two domains: temporal and spatial. Thus, after extracting frame-level spatial features through the PSFE, additional temporal feature modeling is necessary to capture dynamic characteristics.

### *2.3.1 Multi-range temporal modeling*

Feature modeling in the temporal dimension is often susceptible to long-range dependency issues. To enhance the expressiveness of these features, we employ a multi-scale approach to capture features at distinct scales. Specifically, convolutional kernels of varying sizes are utilized to convolve features within the temporal domain. As convolutional kernels of different sizes correspond to different receptive fields, they influence the convolution operation's perception range uniquely. Larger convolutional kernels yield a broader receptive field, which helps mitigate long-range dependency issues. Accordingly, we apply convolution operations across two scales to capture features at varying temporal resolutions, as shown below:

$$\text{Temporal}_{\text{longrange}} = \text{Conv}_{\text{l}}(x_i) \quad (7)$$

$$\text{Temporal}_{\text{shortrange}} = \text{Conv}_{\text{s}}(x_i) \quad (8)$$

where $x_i$ refers to feature map sequence, $i \in 1,2,\cdots,T$, $T$ denotes the number of frames.

*2.3.2 Feature aggregator*

This module weighted vector of global-local feature for aggregation. As shown in Fig. 4, feature aggregator consists of two parallel branches, the part-score branch and the weighted function branch. The former branch further models each spatio-temporal feature vector to highlight the most discriminative spatio-temporal features, specifically for each feature, the vector calculates the importance score through the attention mechanism, and the specific operation is shown as follows:

$$\text{PartScore} = \text{Sigmoid}(\text{Conv1d}(\text{Gelu}(\text{Conv1d}(x_i)))) \quad (9)$$

In the Part-Selection module, the original feature vector is mapped to a scoring feature vector using a max function operation. This transformation can be formally expressed by the following equation:

$$\text{ScoringVector} = \text{Max}(\text{PartScore}(x_i)) \quad (10)$$

The weighed function is used to aggregate prominent temporal features and process the features to eliminate possible temporal domain aliasing. The specific operation of weighted function is shown as follows:

$$\text{WeightedFunction} = \text{AvgPool}() + \text{MaxPool}() \quad (11)$$

Through the above formula, we obtain the weight vector for calculation. The decisive spatio-temporal feature vector obtained through the dot product operation of the weighted vector and the scoring vector is shown as follows:

$$\text{DecisiveVector} = \text{WeightedVector} \cdot \text{PartVector} \quad (12)$$

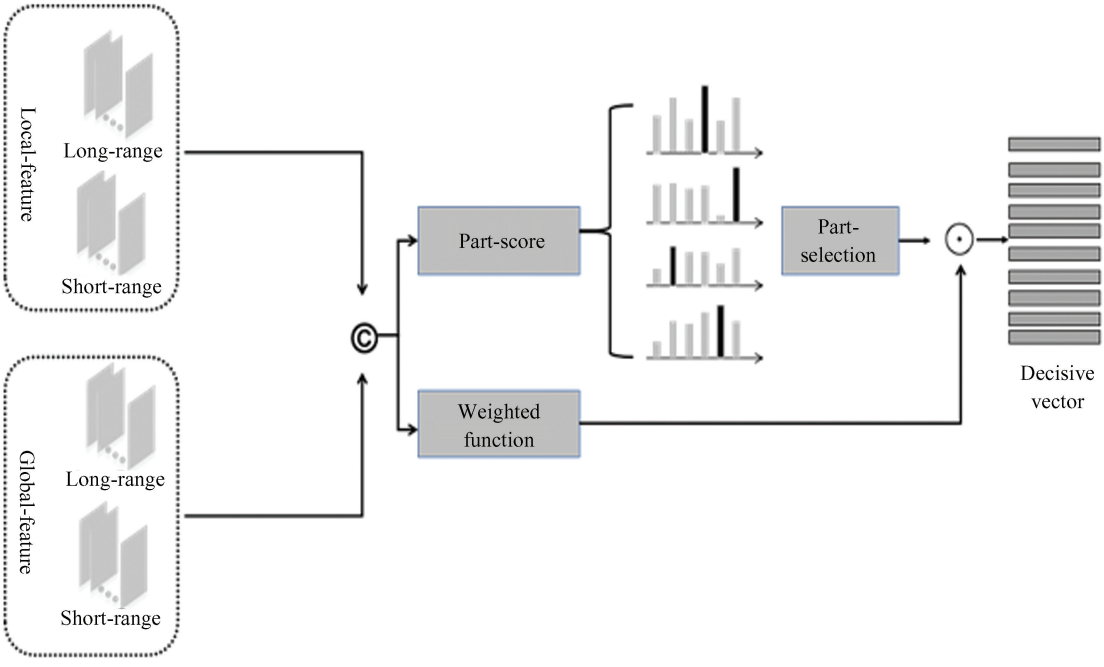



**Fig. 4 The abstract structure of dynamic feature aggregator in practice**

## 3 Experiments

A large number of experiments were conducted on two commonly used datasets to evaluate the proposed algorithm in this section. The relevant details of the CASIA-B dataset[25] are introduced in Section 3.1. In Section 3.2, the experimental settings of the proposed algorithm are described, and in Section 3.3, a comparison of experimental results between the proposed method and current state-of-the-art methods on the CASIA-B dataset[25] is presented.

### 3.1 Datasets

As a popular benchmark dataset for evaluating gait recognition, CASIA-B[25] contained 124 subjects' gait sequences. The subject (from 001 to 124) has 110 sequences totally, which are captured from 11 different viewpoints gaps 18°(0° to 180°). For every viewpoint, 10 sequences are collected under three scenarios: six sequences with normal walking, two sequences carrying a backpack, and two sequences wearing a coat.

We follow dataset partitioning strategies from Refs.[3] and [2], dividing CASIA-B into Large-sample training (LT) setups, Medium-sample training (MT) setups, Small-sample training (ST) setups, with specified training/testing splits, using NM (#01 - 04) sequences as gallery sets and NM (#05 - 06), BG (#01 - 02), CL (#01 - 02) sequences as probe subsets.

### 3.2 Implementation Details

Experiments were conducted on NVIDIA 3090 GPUs using silhouette alignment, Adam optimizer (learning rate $10^{-4}$, momentum 0.2), batch-all separate triplet loss (margin 0.2), and a batch size of (8,12) trained for 100000 iterations.

### 3.3 Comparison with State-of-the-Art Methods

To assess the performance of our approach under cross-view conditions, we conducted a detailed comparison with several leading methods. As presented in Table 1, our method achieves higher recognition accuracy across most viewpoints compared with the seven reference methods. Specifically, for NM/BG/CL conditions, the average recognition accuracy of our method surpasses that of CaitSlice[8] by 0.6%, 1.7%, and 2.3%, respectively. Although our proposed global-local part-shift network significantly improves upon previous single-modal gait recognition methods, it achieves slightly lower average recognition accuracy compared with recent multi-modal approaches such as AttnGait[26] and GMSN[27], specifically 1.9% and 4% lower, respectively. This performance gap is understandable given that multi-modal methods leverage complementary information from multiple modalities (e.g., appearance, pose, or depth data), naturally granting them higher discriminative capabilities.

However, the method proposed in this paper utilizes only single-modal gait silhouettes and still achieves competitive accuracy. Moreover, our approach introduces notable improvements in modeling internal relationships among non-adjacent body parts and effectively addresses long-range temporal dependencies within gait sequences. These enhancements offer substantial value, especially in scenarios where multi-modal data collection may be constrained by privacy concerns, deployment costs, or hardware limitations. Future research could explore integrating our global-local part-shift network with complementary modalities to further bridge this performance gap.

**Table 1 Rank-1 accuracy averaged across three experimental settings on CASIA-B, excluding same-view comparisons** (%)

| Gallery NM#1-4 | | | Accuracy | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CASIA-B | Probe subsets | Methods | 0° | 18° | 36° | 54° | 72° | 90° | 108° | 126° | 144° | 162° | 180° | Mean |
| LT | NM(#5-6) | GaitSet[2] | 90.8 | 97.9 | 99.4 | 96.9 | 93.6 | 91.7 | 95.0 | 97.8 | 98.9 | 96.8 | 85.8 | 95.00 |
| | | GaitSlice[8] | 95.5 | 99.2 | 99.6 | 99.0 | 94.4 | 92.5 | 95.0 | 98.1 | 99.7 | 98.3 | 92.9 | 96.70 |
| | | GaitPart[3] | 94.1 | 98.6 | 99.3 | 98.5 | 94.0 | 92.3 | 95.9 | 98.4 | 99.2 | 97.8 | 90.4 | 96.20 |
| | | AttenGait[26] | 98.4 | 98.2 | 98.6 | 98.1 | 98.7 | 99.0 | 98.9 | 99.2 | 99.3 | 99.3 | 99.3 | 98.80 |
| | | GMSN[27] | 96.7 | 99.1 | 99.4 | 98.3 | 97.3 | 96.5 | 98.4 | 99.0 | 99.4 | 99.2 | 96.4 | 98.20 |
| | | RPNet[28] | 95.1 | 99.0 | 99.1 | 98.3 | 95.7 | 93.6 | 95.9 | 98.3 | 98.6 | 97.7 | 90.8 | 96.60 |
| | | GPGait[29] | – | – | – | – | – | – | – | – | – | – | – | 94.72 |
| | | Our method | **93.7** | **99.0** | **99.2** | **98.1** | **97.0** | **95.3** | **97.6** | **99.1** | **99.3** | **99.1** | **92.4** | **97.30** |
| | BG(#1-2) | GaitSet[2] | 83.8 | 91.2 | 91.8 | 88.8 | 83.3 | 81.0 | 84.1 | 90.0 | 92.2 | 94.4 | 79.0 | 87.20 |
| | | GaitSlice[8] | 90.2 | 96.4 | 96.1 | 94.9 | 89.3 | 85.0 | 90.9 | 94.5 | 96.3 | 95.0 | 88.1 | 92.40 |
| | | GaitPart[3] | 89.1 | 94.8 | 96.7 | 95.1 | 88.3 | 84.9 | 89.0 | 93.5 | 96.1 | 93.8 | 85.8 | 91.50 |
| | | AttenGait[26] | 96.9 | 97.9 | 98.3 | 98.2 | 97.2 | 95.4 | 97.2 | 98.7 | 98.8 | 98.6 | 97.4 | 97.70 |
| | | GMSN[27] | 95.3 | 98.1 | 98.3 | 96..9 | 94.6 | 91.2 | 94.1 | 97.1 | 98.3 | 97.4 | 94.9 | 96.00 |
| | | RPNet[28] | 92.3 | 96.6 | 96.6 | 94.5 | 91.9 | 87.6 | 90.7 | 94.7 | 96.0 | 93.9 | 86.1 | 92.80 |
| | | GPGait[29] | – | – | – | – | – | – | – | – | – | – | – | 89.29 |
| | | Our method | **90.8** | **97.7** | **98.2** | **96.7** | **92.4** | **88.3** | **92.1** | **95.7** | **97.6** | **96.1** | **89.4** | **94.10** |
| | CL(#1-2) | GaitSet[2] | 61.4 | 75.4 | 80.7 | 77.3 | 72.1 | 70.1 | 71.5 | 73.5 | 73.5 | 68.4 | 50.0 | 70.40 |
| | | GaitSlice[8] | 75.6 | 87.0 | 88.9 | 86.5 | 80.5 | 77.5 | 79.1 | 84.0 | 84.8 | 83.6 | 70.1 | 81.60 |
| | | GaitPart[3] | 70.7 | 85.5 | 86.9 | 83.3 | 77.1 | 72.5 | 76.9 | 82.2 | 83.8 | 80.2 | 66.5 | 78.70 |
| | | AttenGait[26] | 91.1 | 95.3 | 96.0 | 95.3 | 89.9 | 88.4 | 89.5 | 91.3 | 88.8 | 89.4 | 86.0 | 91.00 |
| | | GMSN[27] | 80.1 | 92.0 | 94.2 | 90.5 | 85.9 | 80.3 | 84.6 | 90.2 | 92.3 | 89.9 | 77.7 | 87.00 |
| | | RPNet[28] | 75.6 | 87.1 | 88.3 | 83.1 | 78.8 | 78.0 | 79.9 | 82.7 | 83.9 | 78.9 | 66.6 | 80.30 |
| | | GPGait[29] | – | – | – | – | – | – | – | – | – | – | – | 86.65 |
| | | Our method | **76.3** | **89.3** | **91.3** | **87.1** | **81.8** | **77.6** | **83.7** | **87.8** | **89.0** | **85.6** | **73.6** | **83.90** |

Additionally, we evaluated the performance of our approach in data-limited scenarios, with experimental results illustrated in Fig. 5. These results indicate that our method significantly outperforms both GaitSlice[8] and MT3D[30]. In particular, the average recognition accuracy under ST and MT settings is 1.8% and 1.4% higher than GaitSlice[8], and 0.8% and 1.1% higher than MT3D[30], respectively. The superior accuracy under MT further underscores the robustness and efficiency of our method compared with GaitSlice[8].

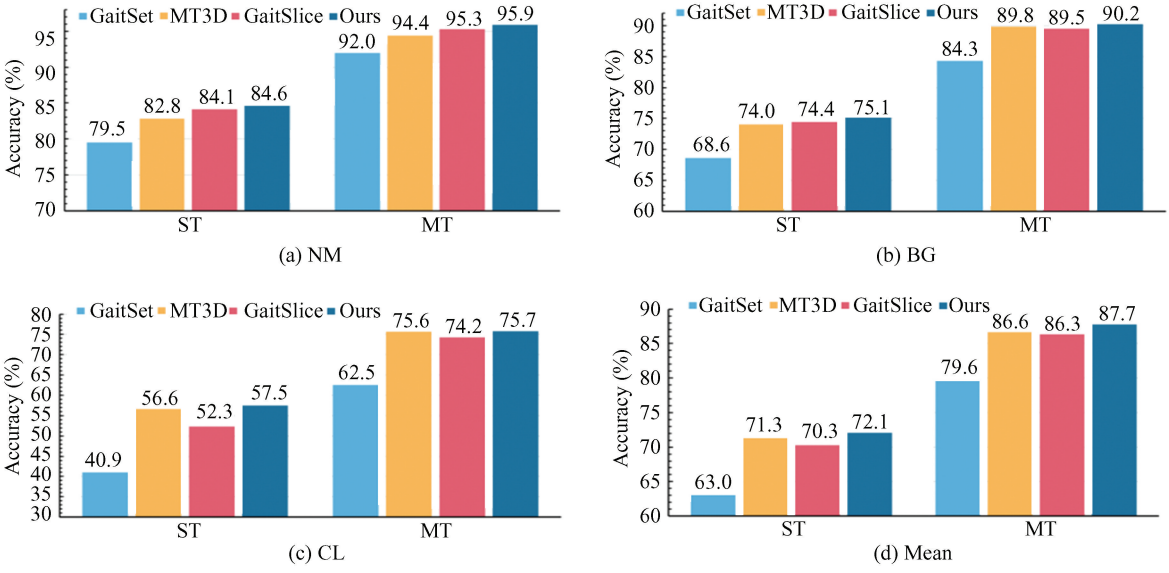



**Fig. 5 Benchmarking against leading approaches under ST/MT configurations**

## 4 Conclusions

In this work, we introduce a distinct global-local representation learning framework that leverages a part-shift mechanism specifically designed to overcome constraints from insufficient visual information and external interferences in gait recognition. The proposed framework comprises two specialized modules: the PSFE and the DFA. Specifically, the PSFE strategically shifts and recombines posture segments to capture intrinsic connections between non-adjacent body parts, thereby significantly enriching both global and local spatial features. Concurrently, the DFA employs multi-range temporal modeling, effectively preserving short-range spatio-temporal patterns and addressing long-range temporal dependencies. Extensive experiments conducted on the widely-used CASIA-B dataset confirm the effectiveness and competitive performance of our method, highlighting its practical applicability and robustness against challenging variations.

## References

[1] He Y, Zhang J, Shan H, et al. Multi-task gans for view-specific feature learning in gait recognition. IEEE Transactions on Information Forensics and Security, 2019, 14(1): 102-113. DOI: 10.1109/TIFS.2018.2844819.

[2] Chao H, He Y, Zhang J, et al. Gaitset: Regarding gait as a set for cross-view gait recognition. Proceedings of the 33rd AAAI Conference on Artificial Intelligence, AAAI 2019, 31st Innovative Applications of Artificial Intelligence Conference, IAAI 2019 and the 9th AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI. Hong Kong, China: AAAI Press, 2019: 8126-8133. DOI: 10.1609/aaai.v33i01.33018126.

[3] Fan C, Peng Y, Cao C, et al. Gaitpart: Temporal part-based model for gait recognition. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 14225 - 14233. DOI: 10.1109/CVPR42600.2020.01423.

[4] Li X, Makihara Y, Xu C, et al. End-to-end model-based gait recognition. Proceedings of the Asian Conference on Computer Vision. Cham: Springer, 2020, 12642: 3 - 20. DOI: 10.1007/978-3-030-69535-4_1.

[5] Liao R, Cao C, Garcia E B, et al. Pose-based temporal-spatial network (PTSN) for gait recognition with carrying and clothing variations. Biometric Recognition: 12th Chinese Conference, CCBR 2017. Cham: Springer, 2017: 474-483. DOI: 10.1007/978-3-319-69923-3_51.

[6] Teepe T, Khan A, Gilg J, et al. Gaitgraph: Graph convolutional network for skeleton-based gait recognition. Proceedings of the 2021 IEEE International Conference on

Image Processing (ICIP). Piscataway: IEEE, 2021:2314-2318. DOI:10.48550/arXiv.2101.11228.

[7] Hou S, Cao C, Liu X, et al. Gait lateral network: Learning discriminative and compact representations for gait recognition. Computer Vision-ECCV 2020. ECCV 2020. Lecture Notes in Computer Science. Cham:Springer, 2020,12354:382-398.DOI:10.1007/978-3-030-58545-7_22.

[8] Li H,Qiu Y, Zhao H, et al. Gaitslice: A gait recognition model based on spatio-temporal slice features. Pattern Recognition, 2022,124:108453. DOI:1016/j.patcog.2021.108453.

[9] Wan J, Zhao H, Li R, et al. Omni-domain feature extraction method for gait recognition. Mathematics, 2023, 11(12):2612. DOI:10.3390/math11122612.

[10] Aung S T Y, Kusakunniran W. A comprehensive review of gait analysis using deep learning approaches in criminal investigation. PeerJ Computer Science, 2024: e2456. DOI: 10.7717/peerj-cs.2456.

[11] Xiao J, Yang H, Xie K, et al. Learning discriminative representation with global and fine-grained features for cross-view gait recognition. CAAI Transactions on Intelligence Technology,2022,7(2):187-199.DOI:10.1049/cit2.12051.

[12] Wang Z, Hou S, Zhang M, et al. LandmarkGait: Intrinsic human parsing for gait recognition. In Proceedings of the 31st ACM International Conference on Multimedia. New York: ACM, 2023: 2305 - 2314. DOI: 10. 1145/3581783.3611840.

[13] Huo W, Wang K, Tang J, et al. GaitSCM: Causal representation learning for gait recognition. Computer Vision and Image Understanding, 2024, 243: 103995. DOI:10.1016/j.cviu.2024.103995.

[14] Feng Y, Yuan J, Fan L. GaitFusion: Exploring the fusion of silhouettes and optical flow for gait recognition. International Conference on Artificial Neural Networks. Cham: Springer Nature Switzerland,2023:88-99.

[15] Wang L, Chen J, Liu Y.Frame-level refinement networks for skeleton-based gait recognition. Computer Vision and Image Understanding, 2022,222:103500. DOI:10.1016/j.cviu.2022.103500.

[16] Zhang Z, Wei S, Xi L, et al. GaitMGL: Multi-scale temporal dimension and global-local feature fusion for gait recognition. Electronics, 2024, 13 (2): 257. DOI: 10. 3390/electronics13020257.

[17] Han J, Bhanu B. Individual recognition using gait energy image. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 28 (2): 316 - 322. DOI: 10. 1109/TPAMI.2006.38.

[18] Bashir K, Xiang T, Gong, S. Gait recognition using gait entropy image. In 3rd International Conference on Imaging for Crime Detection and Prevention. Piscataway: IEEE, 2009:1-6.DOI:10.1049/ic.2009.0230.

[19] Wang K, Liu L, Lee Y, et al., 2019. Nonstandard periodic gait energy image for gait recognition and data augmentation. Pattern Recognition and Computer Vision: Second Chinese Conference, PRCV 2019. Cham: Springer, 2019: 197 - 208. DOI: 10. 1007/978 - 3 - 030 - 31723-2_17.

[20] Xiong J, Zou S, Tang J. DFGait: Decomposition fusion representation learning for multimodal gait recognition. International Conference on Multimedia Modeling. Cham: Springer Nature Switzerland, 2024: 381 - 395. DOI: 10. 1007/978-3-031-53311-2_28.

[21] Bai S, Chang H, Ma B. Incorporating texture and silhouette for video-based person re-identification. Pattern Recognition, 2024, 156: 110759. DOI: 10. 1016/j. patcog. 2024.110759.

[22] Chen B, Niu T, Yu W, et al. A-net: An a-shape lightweight neural network for real-time surface defect segmentation. IEEE Transactions on Instrumentation and Measurement, 2023,73: 1-14.DOI:10.1109/TIM.2023.3341115.

[23] Cui C, Liu L, Qiao R. A cutting-edge video anomaly detection method using image quality assessment and attention mechanism-based deep learning. Alexandria Engineering Journal, 2024,108:476-485. DOI:10.1016/j.aej.2024.07.103.

[24] Liu F,Wang Q, Xiao Y, et al. 2025. An efficient and effective pore matching method using ResCNN descriptor and local outliers. Pattern Recognition,2025,163:111446. DOI:10.1016/j.patcog.2025.111446.

[25] Yu S, Tan D, Tan T. A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06). Piscataway: IEEE, 2006:441-444. DOI: 10.1109/ICPR.2006.67.

[26] Castro F M, Delgado-Escaño R, Hernández-García R, et al. AttenGait: Gait recognition with attention and rich modalities. Pattern Recognition, 2024,148:110171. DOI: 10.1016/j.patcog.2023.110171.

[27] Wei T, Liu M, Zhao H, et al. GMSN: An efficient multi-scale feature extraction network for gait recognition. Expert Systems with Applications, 2024, 252: 124250. DOI:10.1016/j.eswa.2024.124250.

[28] Qin H, Chen Z, Guo Q, et al. RPNet: Gait recog-nition with relationships between each body-parts. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 32: 2990 - 3000. DOI: 10. 1109/TCSVT.2021.3095290.

[29] Fu Y, Meng S, Hou S, et al. GPGait: Generalized pose-based gait recognition. Proceedings of the IEEE/CVF International Conference on Computer Vision. Paris: CVF, 2023:19595-19604.

[30] Lin B, Zhang S, Bao F. Gait recognition with multiple-temporal-scale 3d convolutional neural network. Proceedings of the 28th ACM International Conference on Multimedia. New York: ACM, 2020: 3054 - 3062. DOI: 10. 1145/3394171.3413861.