Brazilian Journal of Biometrics, **43**, 1–24 (2025)
doi:10.28951/bjb.v43i4.857

BRAZILIAN JOURNAL OF
BIOMΣTRICS
ISSN:2764-5290

ARTICLE

Optimizing Gait-Based Biometric Recognition with Deep Dense Networks

Sachin Bhimraj Mandlik,* Rekha Punjaji Labade, Sachin Vasant Chaudhari, and Balasaheb Shrirangao Agarkar

Department of Electronics and Telecommunication Engineering, Sanjivani College of Engineering, Kopargaon, India, Savitribai Phule Pune University, Pune, India
*Corresponding author. Email: mandlik.sb@gmail.com

(Received: March 10, 2025; Revised: July 17, 2025; Accepted: August 22, 2025; Published: November 25, 2025)

Abstract

Gait recognition is a developing biometric technique capable of identifying individuals from a distance, with wide-ranging applications such as video surveillance. A primary challenge is the extraction of discriminative gait features from silhouettes that are robust to variations in apparel, carried objects, and camera viewpoints. To address these limitations, this study introduces GaitSTR — a novel framework that harnesses pyramid mapping for enhanced temporal and spatial feature extraction, integrated with a deep neural network comprising dense layers. Pyramid mapping decomposes gait sequences into multi-scale spatial features, enabling GaitSTR to capture fine-to-coarse motion patterns and improve recognition under varying conditions. The method focuses on extracting distinctive feature representations at different frame levels, effectively utilizing spatial and temporal variations within video sequences. The proposed model utilizes a memory-augmented recurrent neural network (RNN) enriched with temporal attention to capture sequential motion cues, while spatial features are extracted through a densely connected attention-guided network By employing the pyramid-based hierarchical feature extraction, along with attention mechanisms in both spatial and temporal component, the network can prioritize the most significant video segments, improving its efficiency and learning capacity for processing intricate gait data. The results are evaluated on four widely used benchmark datasets: GREW, OU-ISIR, OU-MVLP, and CASIA-B—achieving 92.4% on GREW, 95.2% on OU-ISIR, and 0.96 mean accuracy on OU-MVLP, and 98.4% (normal) on CASIA-B, surpassing state-of-the-art methods. These results underscore the robustness of our approach under diverse conditions, establishing a new benchmark for performance in gait recognition.

Keywords: Deep Learning; Pyramid mapping; Spatial variation; Time-based modulation; Gait Recognition.

1. Introduction

Gait recognition has emerged as a prominent biometric technique for identifying individuals based on their unique walking patterns. Gait offers the distinct advantage of enabling identification from a distance without requiring active cooperation from the subject (Sethi *et al.,* 2022; Song *et al.,* 2024). This contrasts with traditional anatomical biometrics such as fingerprints, facial features, DNA, or iris patterns (Choi *et al.,* 2019; Gadaleta & Rossi, 2018; Ma *et al.,* 2017). These features have made gait recognition a prominent area of research for surveillance-related applications, such as, biometric criminal investigation, law enforcement civil security, and smart transportation systems (Balazia & Sojka, 2018; Bastos & Tavares, 2025; Yao *et al.,* 2022). As a result, significant research attention has been devoted to gait recognition in recent years(Sepas-Moghaddam & Etemad, 2023). Recent advancements in technology have established gait analysis (Cai *et al.,* 2023; Prajapati *et al.,* 2021) as a reliable and non-invasive method for clinical evaluation, particularly in the diagnosis of health conditions, identification of individuals, and assessment of locomotor patterns (Erdaş *et al.,* 2021; Panahi & Ghods, 2018). However, the accuracy of gait recognition is often challenged by external variations such as changes in attire, carried objects, and camera perspectives (Hou *et al.,* 2023; Huo *et al.,* 2026; Mandlik *et al.,* 2025b). This underscores the necessity of improving model robustness under diverse and unconstrained conditions (Wei *et al.,* 2024). Earlier studies have introduced multiple strategies to mitigate the challenges posed by changes in viewpoint, clothing, and carried objects (Gao *et al.,* 2022; Mitra & Acharya, 2007; Xu *et al.,* 2021). These traditional approaches are typically categorized into two groups: those utilizing the Gait Energy Image (GEI) (Ben *et al.,* 2020; Chen *et al.,* 2018; Gupta & Chattopadhyay, 2021; Huang & Boulgouris, 2012; Mogan *et al.,* 2024) and those that interpret gait as sequence-independent sets (Chao *et al.,* 2022; Lin *et al.,* 2021). While GEI-based techniques have been widely used, they often fail to capture detailed spatiotemporal cues, which can adversely affect recognition accuracy. In contrast, methods that represent gait as sequence-independent sets—though they have yielded promising results in previous studies—have primarily demonstrated effectiveness within controlled laboratory environments. To mitigate the aforementioned challenges—such as variations in clothing, carried items, and camera viewpoints—this study introduces GaitSTR for robust gait recognition. The approach combines pyramid mapping with a densely layered deep neural architecture to improve the capture of motion patterns across both spatial scales and time. By breaking down gait sequences into multiple resolution levels, pyramid mapping facilitates the extraction of motion cues ranging from detailed to broader movements, thereby increasing resilience to visual inconsistencies. To model time-based dynamics effectively, the framework incorporates a memory-augmented RNN with a time-based attention mechanism, enabling it to concentrate on crucial frame-level information. Simultaneously, a densely connected convolutional network embedded with channel and spatial attention modules enhances spatial feature discrimination by adapting to variations across different channels and regions. This integrated strategy—leveraging hierarchical decomposition and attention-based refinement—allows GaitSTR to emphasize the most informative portions of a video sequence, leading to more accurate and robust gait feature learning.

2. Literature Survey

Over the past three decades, research in Gait Recognition field has evolved from early model-based approaches to sophisticated deep learning techniques, significantly improving recognition accuracy and robustness. This literature review provides a comprehensive examination of gait recognition methodologies.

2.1 Traditional Gait Recognition Approaches

The conceptual foundation of gait recognition can be traced back to psychological studies on human motion perception. The seminal work (Johansson, 1973) demonstrated that individuals could recognize biological motion using only point-light displays, suggesting that gait contains distinctive patterns that can be computationally modeled. This discovery inspired early research efforts in the 1990s to develop automated gait recognition systems. A model-based approach (Niyogi & Adelson, 1994) was among the first proposed, analyzing motion trajectories to extract gait characteristics. Their work established that kinematic features, such as stride length and joint angles, could be used for identification. Fourier analysis was applied to model leg movements, demonstrating that gait could be represented mathematically for recognition purposes (Cunado *et al.,* 2003). These pioneering studies confirmed the viability of gait as a biometric trait and set the stage for more structured research in the early 2000s. Early gait recognition systems primarily employed two methodological paradigms: model-based and appearance-based techniques. Model-based approaches relied on constructing biomechanical representations of the human body to extract gait-related features. A stride-based model was developed to measure step length and walking speed, achieving reasonable accuracy in controlled environments (BenAbdelkader *et al.,* 2002).However, this approach was sensitive to variations in walking speed and camera angles. A kinematic model was later introduced to track hip and knee movements, improving robustness against minor viewpoint changes (Bouchrika & Nixon, 2008). Despite their interpretability, model-based methods faced significant challenges due to their dependency on accurate pose estimation, which was difficult to achieve with low-resolution or occluded video footage. Appearance-based methods, in contrast, avoided explicit modeling by analyzing the silhouette of a walking person. These approaches gained popularity due to their computational efficiency and effectiveness in controlled settings. A significant contribution was the introduction of the Gait Energy Image (GEI), a compact representation that averaged silhouette sequences over a gait cycle (Han & Bhanu, 2006). The GEI became a benchmark for subsequent research due to its ability to capture temporal gait dynamics in a single image. Silhouette-based recognition was further enhanced by employing Dynamic Time Warping (DTW) to align gait sequences temporally, addressing variations in walking speed (Liu & Sarkar, 2006). Despite their advantages, appearance-based methods were sensitive to changes in clothing, carrying conditions (e.g., backpacks or bags), and camera viewpoints, limiting their real-world applicability (Chao *et al.,* 2022; Liu *et al.,* 2021; Mandlik *et al.,* 2025c; Zou *et al.,* 2025).

2.2 Deep Learning Revolution in Gait Recognition

The advent of deep learning in the 2010s brought transformative changes to gait recognition, enabling end-to-end learning of discriminative features from raw data. Convolutional Neural Networks (CNNs) emerged as the dominant architecture for gait analysis due to their ability to learn hierarchical spatial representations. CNNs were among the first applied to GEI, demonstrating superior performance compared to traditional methods (Tong *et al.,* 2017). A paradigm shift was later introduced with GaitSet, which treated gait as an unordered set of silhouettes rather than a fixed sequence. This approach significantly improved cross-view recognition by eliminating the need for strict temporal alignment (Chao *et al.,* 2022). The success of GaitSet highlighted the potential of set-based representations in handling variable gait cycle lengths and occlusions. Recognizing that gait is inherently a spatio-temporal process, researchers began incorporating recurrent architectures and 3D convolutional networks to better capture motion dynamics. One approach combined 3D CNNs with Long Short-Term Memory (LSTM) networks to model both spatial and temporal gait features, achieving robust performance across different walking speeds (Liu *et al.,* 2019). Further advancement came with GaitPart, which focused on fine-grained part-level features to improve recognition under varying carrying conditions (Fan *et al.,* 2020). GaitPart's emphasis on local temporal dynamics demonstrated that part-based approaches could enhance robustness against appearance variations. Most recently, transformer-based architectures have been applied to gait recognition, leveraging self-attention mechanisms to capture long-range dependencies in gait sequences. GaitGL integrated global and local features using transformer modules, achieving state-of-the-art performance on benchmark datasets (Lin *et al.,* 2021). A pure transformer-based model, GaitFormer, was introduced and outperformed CNN-based methods in cross-view recognition tasks (Li *et al.,* 2024a). These advancements underscore the growing influence of transformer architectures in gait recognition, particularly in handling complex variations in viewpoint and appearance. Despite significant progress, gait recognition systems still face several challenges that hinder their deployment in real-world scenarios. Viewpoint variation remains a critical issue, as most systems experience performance degradation when the camera angle changes. Recent work addressed this through adversarial learning, generating view-invariant gait representations (He *et al.,* 2019). Clothing and carrying conditions continue to pose difficulties, as heavy coats or bags can alter gait appearance. A domain adaptation network called GaitDAN improved robustness against such variations by aligning feature distributions across different domains (Huang *et al.,* 2024). Occlusion and low-resolution data present additional challenges, particularly in surveillance applications where subjects may be partially obscured. One solution uses attention mechanisms to focus on visible body parts while reconstructing missing information (Hasan *et al.,* 2024). Cross-domain generalization is another persistent issue, as models trained in laboratory settings often fail in real-world environments. Unsupervised and self-supervised learning approaches, aim to bridge this gap by reducing reliance on labeled data (Pinčić *et al.,* 2022; Wang *et al.,* 2025b). Recurrent Neural Networks (RNNs) (Rashmi & Guddeti, 2022; Xing *et al.,* 2018; Zhang *et al.,* 2022) capture temporal dependencies, while Deep Autoencoders (DAe) (Li *et al.,* 2019; Song *et al.,* 2019) learn compact gait representations. Hybrid models combining CNNs, RNNs, and DAe further improve recognition accuracy by leveraging their complementary strengths (Zhang *et al.,* 2020; Zhang *et al.,* 2022).

While recent gait recognition methods have shown considerable promise, they often encounter difficulties in capturing distinctive motion features under real-world challenges such as changes in attire (Altab Hossain *et al.,* 2010; Castro *et al.,* 2024) , carried objects (Mizuno *et al.,* 2024; Uddin *et al.,* 2018), and varying camera viewpoints (Du & Zhao, 2024; Makihara *et al.,* 2015; Muramatsu *et al.,* 2015). Moreover, their non-end-to-end architectures—typically involving separate stages for 3D reconstruction, feature extraction, and gait matching—introduce significant limitations by increasing computational complexity and reducing overall efficiency for practical deployment (Filipi Gonçalves Dos Santos *et al.,* 2023; Hasan *et al.,* 2024; Li *et al.,* 2024b; Mandlik *et al.,* 2025a; Sepas-Moghaddam & Etemad, 2023; Sokolova & Konushin, 2019). To overcome these obstacles, this study introduces GaitSTR, a unique framework employing pyramid mapping to perform hierarchical spatial and temporal feature extraction. By decomposing gait sequences into multiple spatial scales, the method captures motion patterns ranging from fine to coarse granularity. The framework integrates a deep densely connected network to extract spatial features and a memory-augmented recurrent neural network with temporal attention to capture sequential dependencies in motion. This combined strategy allows the model to focus on the most informative segments of the gait cycle, reducing noise and improving recognition performance across varied conditions, thereby enhancing its suitability for real-world applications.

3. Materials and Methods

The block diagram of the proposed GaitSTR framework, illustrated in Figure 1, presents a structured flow for processing gait sequences and extracting discriminative spatio-temporal features. The process begins with input silhouettes derived from a gait video sequence that captures the walking motion of an individual. These silhouettes are first processed by the Pyramid Mapping Module, which decomposes the input into fine-to-coarse spatial scales to enhance multi-level motion representation. The resulting features are then passed through two parallel processing streams: one stream is input to an Attention-Guided Network to extract salient spatial features using temporal attention, while the other stream is processed by a Memory-Augmented Recurrent Neural Network (RNN) to capture sequential motion cues across frames. The outputs of these spatial and temporal pathways are integrated in the final Gait Prediction block, which implicitly performs feature fusion by combining spatial attention and sequential dependencies, followed by classification to identify the subject based on the learned gait features.

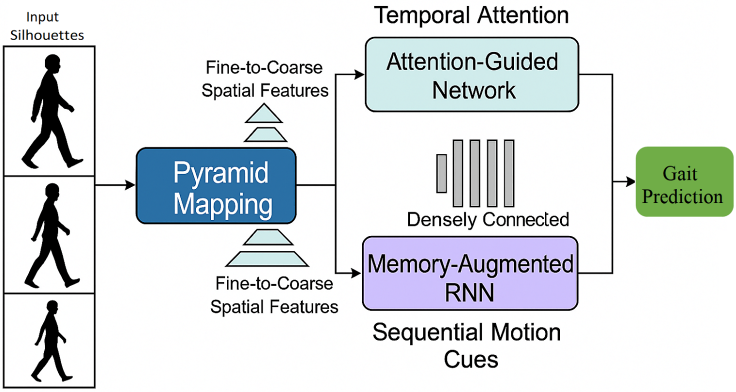

Figure 1. The GaitSTR framework.

3.1 Frame-wise Detection and ROI extraction

As illustrated in Figure 2, the system performs object detection on each input video frame, focusing on identifying and isolating human subjects. For each detected individual, bounding boxes are generated dynamically to localize the region of interest, ensuring that the gait-specific features are captured accurately and consistently across frames. This process significantly reduces background noise and irrelevant visual data, leading to more precise silhouette extraction and improved recognition performance. Following object detection and ROI extraction, the system proceeds with detailed preprocessing and enhancement of the extracted gait silhouettes.

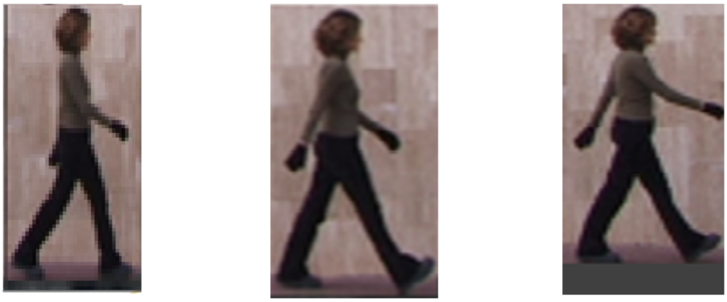

Figure 2. Object detection is performed on each frame, and bounding boxes are generated around detected objects.

3.2 Data Augementation

Data augmentation is a technique used to enhance the size and diversity of a dataset by introducing random transformations to the original data. For instance, images can be rotated, cropped, or flipped. This approach is commonly employed to reduce the risk of overfitting, ensuring the model's robustness and its ability to generalize effectively during the training process. The Table 1 below outlines the parameters employed for data augmentation during the training process. Figure 3 provides a subjective depiction of the results produced by data augmentation.

Table 1. Augmentation Type and parameters

Serial Number	Augmentation Type	Parameter Details
1	Rotation	Random rotation within -10 to 10 degrees
2	Reflection (X-axis)	No specific variation applied
3	X-axis Translation	Random shift in the range of -5 to 5 pixels
4	Y-axis Translation	Random shift in the range of -5 to 5 pixels

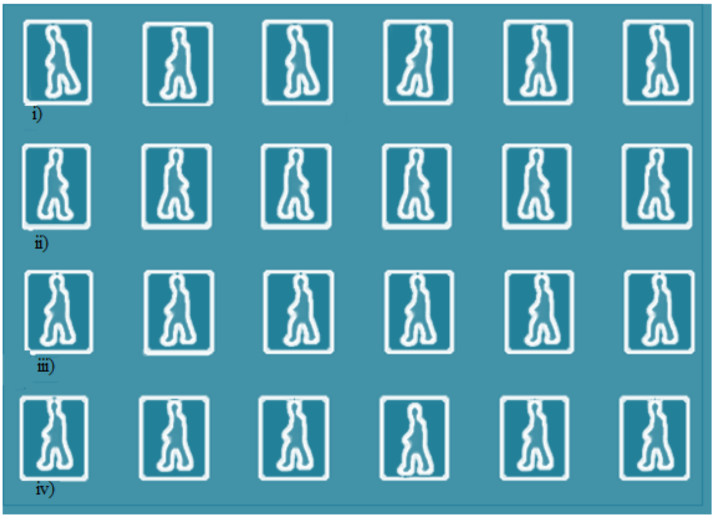

Figure 3. Object detection is performed on each frame, and bounding boxes are generated around detected objects.

3.3 Pyramid Mapping Module

The **Pyramid Mapping Module** is a core component of the proposed GaitSTR framework, designed to enhance the representation of spatial and temporal dynamics in gait sequences. This module decomposes the input silhouette sequence into a hierarchy of spatial scales, such as fine, medium, and coarse resolutions, enabling the model to capture motion information at multiple levels of abstraction.

Given an input silhouette sequence

$$S = \lfloor s_1, \ldots, s_t, \ldots, s_T \rfloor \tag{1}$$

where T denotes the number of frames, shallow spatial features $F_t \in \mathbb{R}^{H \times W \times C}$ are extracted from each frame s_t using a base convolutional feature extractor:

$$F_t = \phi(s_t), \quad t = 1, 2, \ldots, T \tag{2}$$

where $\phi(\cdot)$ denotes the initial feature extraction function.

At finer scales, the module preserves subtle and localized motion cues such as limb movement and foot positioning, while coarser scales retain global structural patterns like posture and walking trajectory. Each feature map F_t is then decomposed into a set of multi-scale representations

$$\{F_t^{(1)}, F_t^{(2)}, \ldots, F_t^{(L)}\} \tag{3}$$

corresponding to L spatial levels.

For scale l, the feature map is down-sampled or partitioned into R_l spatial regions:

$$F_t^{(l)} = \text{Pool}_{R_l}(F_t), \quad l = 1, 2, \ldots, L \tag{4}$$

This multi-resolution decomposition provides robustness against common variations in gait data, including changes in clothing, carried objects, and viewpoint shifts. Each spatial scale acts as a complementary representation, allowing the network to better capture the diversity of gait features across time.

The extracted multi-scale feature maps are then forwarded to the subsequent processing streams for spatial attention modeling and temporal sequence learning. These multi-scale features are concatenated along the spatial axis to form the final pyramid-mapped representation:

$$\tilde{F}_t = \text{Concat}\left(F_t^{(1)}, F_t^{(2)}, \ldots, F_t^{(L)}\right) \tag{5}$$

This aggregated representation $\tilde{F}_t$ is forwarded to the spatial and temporal branches of the network, facilitating robust spatio-temporal learning. By encoding fine-to-coarse spatial information, the Pyramid Mapping Module acts as a powerful feature encoder, ensuring that both detailed and holistic motion patterns are available for downstream processing in the GaitSTR architecture.

3.4 Attention-Guided Network (Spatial Stream)

The Attention-Guided Network is responsible for learning spatially discriminative representations by dynamically identifying and emphasizing the most informative regions in each frame of the gait sequence. After multi-scale decomposition in the Pyramid Mapping Module, the resulting features

$$\tilde{F}_t \in \mathbb{R}^{H \times W \times C} \tag{6}$$

retain both local and global motion patterns. However, not all regions within these feature maps contribute equally to identifying an individual's gait.

To address this, a spatial attention mechanism is introduced. A learnable attention map

$$A_t \in [0, 1]^{H \times W} \tag{7}$$

is generated using a lightweight convolutional layer followed by a sigmoid activation:

$$A_t = \sigma\left(\text{Conv}(\tilde{F}_t)\right) \tag{8}$$

This attention map highlights spatial locations that are most relevant for recognition, such as legs, feet, or torso orientation. The final spatially refined feature map is obtained by applying element-wise multiplication:

$$F_t^s = \tilde{F}_t \odot A_t \tag{9}$$

Here, $\odot$ denotes the Hadamard (element-wise) product. These attended features F_t^s are then aggregated across time or directly passed to the fusion layer, preserving only the most discriminative spatial cues while suppressing irrelevant background information or occlusions. This improves robustness to noise and enhances the generalization capability of the network across varying conditions.

3.5 Memory-Augmented Recurrent Neural Network (Temporal Stream)

The Memory-Augmented RNN module models the temporal evolution of gait patterns by learning long-range dependencies between frames in a sequence. Human gait is inherently periodic, with subtle variations across time that carry unique identity cues. Capturing these sequential dynamics is crucial for effective recognition.

Given the sequence of multi-scale features

$$\{\tilde{F}_1, \tilde{F}_2, \tilde{F}_3, \ldots, \tilde{F}_T\} \tag{10}$$

a recurrent unit processes each time step:

$$h_t = \mathrm{RNN}(\tilde{F}_t, h_{t-1}) \tag{4}$$

where h_t represents the hidden state at time t, carrying accumulated temporal information up to that point.

To further enhance this module, a temporal attention mechanism is applied. Not all frames are equally informative—some contain more distinctive phases of the gait cycle (e.g., leg crossing or foot contact). Attention weights α_t are computed over the hidden states:

$$\alpha_t = \frac{\exp(W_a \cdot h_t)}{\sum_{k=1}^{T} \exp(W_a \cdot h_k)} \tag{5}$$

where $W_a \in \mathbb{R}^{1\times d}$ is a trainable weight vector, $h_t \in \mathbb{R}^d$ is the RNN output at time t, and F^m is the weighted feature capturing the entire motion sequence. These weights are used to compute a weighted sum of all hidden states, resulting in a temporally aggregated feature:

$$F^m = \sum_{t=1}^{T} \alpha_t \cdot h_t \tag{6}$$

The memory-augmented mechanism enhances the RNN's ability to preserve long-term dependencies and focus on key temporal cues, making it particularly effective for modeling complex gait patterns, especially in unconstrained environments.

3.6 Attention Unit

The Attention Unit, as indicated in Figure 4, serves as a crucial component in the GaitSTR framework, allowing the model to dynamically focus on the most salient spatial and temporal features within the gait sequence. This unit enhances interpretability and performance by allocating greater importance to discriminative regions and frames.

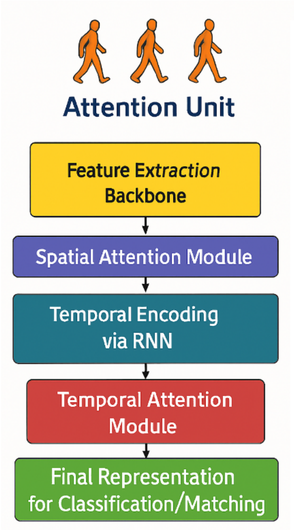
Attention Unit
Feature Extraction Backbone
Spatial Attention Module
Temporal Encoding via RNN
Temporal Attention Module
Final Representation for Classification/Matching

Figure 4. The Attention Unit.

3.7 Spatial Attention Module

Each silhouette frame s_t is first processed to extract a spatial feature map $F_t \in \mathbb{R}^{H\times W\times C}$. However, not all spatial regions within a frame are equally informative—some parts of the body (such as leg movement) often carry more identity cues than others. To selectively emphasize these regions, a spatial attention map $A_t \in [0,1]^{H\times W}$ is learned. The spatial attention map is computed using a lightweight network—typically involving 1 × 1 convolutions and activation functions such as softmax or sigmoid—to produce an attention mask that is broadcast across all channels of the input feature map:

$$\hat{F}_t = A_t \odot F_t \tag{7}$$

where $\hat{F}_t$ is the spatially attended feature, $\odot$ denotes element-wise multiplication, and A_t emphasizes key regions such as limbs or torso depending on their motion saliency.

3.8 Temporal Attention Module

Once the spatial features across time $\{\hat{F}_1, \hat{F}_2, \ldots, \hat{F}_T\}$ are encoded into temporal representations via an RNN, the model generates a sequence of hidden states $\{h_1, h_2, \ldots, h_T\}$. However, different frames in a gait sequence carry varying importance depending on walking phase, occlusion, or noise. To manage this, the temporal attention mechanism learns a weight $\alpha_t \in [0,1]$ for each time step, capturing the relative contribution of each frame toward the final prediction. The equation for the attention weight and the final aggregated temporal feature is given by equations (5) and (6), respectively.

This dual-attention mechanism enables *GaitSTR* to adaptively focus on the most relevant spatial and temporal elements, improving robustness against real-world variations such as occlusion, clothing changes, and speed fluctuations.

4. Classification

One of the key components in the training strategy is a hybrid loss function, which combines Triplet Loss(Wang *et al.*, 2020) and ArcFace Loss (Deng *et al.*, 2019). This dual-loss strategy optimizes the learning of feature embeddings in order to maximize discrimination—an essential requirement for real-world performance since the differences within a class are often far smaller than the similarities between classes.

The Triplet Loss proceeds as follows: triplets are formed with an *anchor* sample, a *positive* sample (same identity), and a *negative* sample (different identity). The objective is to ensure that the distance between the anchor and positive is closer than that between the anchor and negative by at least a margin α. This encourages the network to produce tightly clustered feature-level representations for each identity while maintaining separation from other identities, thereby directly shaping the geometry of the embedding space. The Triplet Loss is defined as:

$$\mathcal{L}_{\text{triplet}} = \sum_{i=1}^{N} \left[|0 f(x_i^a) - f(x_i^p)|0^2 - |0 f(x_i^a) - f(x_i^n)|0^2 + \alpha \right]^+ \tag{11}$$

where $f(x)$ denotes the embedding of sample x, x_i^a, x_i^p, and x_i^n represent the anchor, positive, and negative samples, respectively, $|0 \cdot |0^2$ is the squared Euclidean norm, α is the predefined margin, and $[\cdot]^+$ denotes the hinge function $\max(0, \cdot)$. This loss enforces intra-class compactness and inter-class separation by dynamically adjusting the feature space based on sample relationships.

On the other hand, ArcFace Loss introduces an *angular margin* into the softmax loss, transforming the optimization from Euclidean distance to angular distance (Deng *et al.,* 2019). This imposes a stronger decision boundary by normalizing features onto a hypersphere and penalizing angular deviations from class centers. The ArcFace Loss is expressed as:

$$\mathcal{L}_{\text{arc}} = -\frac{1}{N} \sum_{i=1}^{N} \log \frac{e^{s \cdot \cos(\theta_{y_i} + m)}}{e^{s \cdot \cos(\theta_{y_i} + m)} + \sum_{j \neq y_i} e^{s \cdot \cos(\theta_j)}} \tag{12}$$

where θ_{y_i} is the angle between the embedding vector of the i^{th} sample and its corresponding class center, m is the additive angular margin, s is the scaling factor applied to the feature vectors, and $j \neq y_i$ refers to all classes other than the correct class. The term $\cos(\theta_{y_i} + m)$ increases the angular separation between classes in the feature space.

Combining these two loss functions—Triplet Loss, which promotes intra-class compactness and inter-class separation, and ArcFace Loss, which imposes angular margin constraints—yields a more structured and discriminative embedding space. The total loss is formulated as:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{triplet}} + \mathcal{L}_{\text{arc}} \tag{13}$$

This integrated loss strategy enables the network to learn embeddings that are both tightly clustered within identities and well-separated across different classes. Consequently, the model exhibits enhanced generalization capability across a range of challenging conditions, including variations in walking speed, viewpoint changes, clothing alterations, and partial occlusions, thus improving the robustness and accuracy of gait recognition in real-world scenarios.

5. Experiments

5.1 Datasets

GREW (Guo *et al.,* 2025) includes gait data from 27,345 subjects, split into 20,000 for training (102,887 sequences), 345 for validation (1,784 sequences), and 6,000 for testing (24,000 sequences). Collected in unconstrained environments, it features uncontrolled camera setups, irregular walking paths, and variations in viewpoint, clothing, and occlusion. The number of sequences per subject is limited and random, introducing diverse covariate distributions and class imbalance, making it well-suited for evaluating model generalization in real-world conditions.

OU-ISIR (Iwama *et al.,* 2012) consists of 4,007 subjects (2,135 males and 1,872 females) aged 1–94 years. Gait sequences are captured from four view angles (55°, 65°, 75°, and 85°) under a single walking condition. Each subject has one gallery and one probe sequence.

OU-MVLP (Takemura *et al.,* 2018) is among the largest publicly available gait datasets, containing 10,307 participants recorded from 14 view angles. Each subject has two sequences per view—'Seq-01' for the gallery and 'Seq-00' for probes—supporting large-scale cross-view gait recognition studies.

CASIA-B (Yu *et al.,* 2006) comprises 124 subjects recorded under three conditions: normal walking (NM), walking with a bag (BG), and walking with a coat (CL), across 11 views ranging from 0°–180°. For evaluation, the first 74 subjects form the training set and the remaining 50 the testing set. NM-1 to NM-4 sequences are used as the gallery, while the remaining NM, BG, and CL sequences serve as probes.

5.2 Implementation details

To maintain computational efficiency, all input gait silhouettes are resized to 64×44. Model optimization is carried out using the Adamax optimizer due to its adaptability to complex gait patterns. The mini-group size—representing the number of subjects and sequences per subject—is configured as $(8, 16)$ for the CASIA-B dataset, $(32, 16)$ for both the OU-MVLP and OU-ISIR datasets, and $(32, 8)$ for the GREW dataset. Training is performed with a constant learning rate of 1×10^{-4}, running for 300,000 iterations on CASIA-B, 350,000 iterations on OU-MVLP and OU-ISIR, and 400,000 iterations on GREW.

The GaitSTR model contains 9.8M parameters, requires 2.3 GFLOPs per sequence, and achieves an average inference speed of 12.5 ms per sequence on an NVIDIA RTX 3060 GPU, enabling real-time deployment.

6. Results and Comparative Analysis

The performance of GaitSTR is benchmarked against a diverse set of state-of-the-art gait recognition methods, including recent deep learning as well as classical and hybrid techniques.

6.1 GaitSTR Performance Evaluation against the GREW Dataset

Table 2 presents a comparative analysis of state-of-the-art gait recognition methods on the GREW dataset using Rank-1, Rank-5, Rank-10, and Rank-20 accuracies. Earlier approaches such as GEINet (Shiraga *et al.,* 2016) and GaitPart (Fan *et al.,* 2020) achieve lower recognition rates, while more recent models including DyGait(Wang *et al.,* 2023c), GaitMoE (Huang *et al.,* 2025), and GaitC3I (Wang *et al.,* 2025a) show considerable improvement in performance. The proposed GaitSTR surpasses all existing methods, achieving the highest accuracies across all ranks—83.5% at Rank-1, 91.7% at Rank-5, 86.3% at Rank-10, and 92.4% at Rank-20—setting a new benchmark on the GREW dataset. These results clearly demonstrate the robustness and effectiveness of the GaitSTR framework under diverse and challenging real-world conditions.

The superior performance of GaitSTR can be attributed to its pyramid-based spatial-temporal feature decomposition and refined attention mechanisms. It consistently outperforms attention-guided and expert-mixture networks, validating its discriminative feature learning capabilities. The architecture exhibits strong generalization, maintaining high accuracy across a range of variations such as clothing, speed, and occlusion. Overall, GaitSTR proves to be a reliable and scalable solution for real-world gait recognition applications.

Table 2. Rank-1, Rank-5, Rank-10, and Rank-20 Accuracy on the GREW Dataset.

Method	Rank-1	Rank-5	Rank-10	Rank-20
GaitMPA (Huo *et al.*, 2026)	70.9	83.4	87.5	90.0
GaitGCI (Dou *et al.*, 2023)	68.5	80.8	84.9	87.7
GaitSet (Chao *et al.*, 2022)	46.3	63.6	70.3	76.8
DyGait (Wang *et al.*, 2023c)	71.4	83.2	86.8	89.5
MTSGait (Zheng *et al.*, 2022)	55.3	71.3	76.9	81.6
GaitPart (Fan *et al.*, 2020)	44.0	60.7	67.3	73.5
CSTL (Huang *et al.*, 2021)	50.6	65.9	71.9	76.9
GEINet (Shiraga *et al.*, 2016)	6.8	13.4	17.0	21.0
GaitGL (Lin *et al.*, 2021)	47.3	63.6	69.3	74.2
OpenGait (Fan *et al.*, 2023)	60.1	75.8	–	–
GaitCSV (Wang *et al.*, 2023a)	64.9	78.7	–	–
HSTL (Wang *et al.*, 2023b)	62.7	76.6	–	–
CLASH (Dou *et al.*, 2025)	67.0	78.9	–	–
CLTD (Xiong *et al.*, 2025)	78.0	87.8	–	–
GaitMoE (Huang *et al.*, 2025)	79.6	89.1	–	–
DeepGaitV2 (Wang *et al.*, 2025a)	79.4	88.9	–	–
DeepGaitV2-30 (ibid.)	79.5	–	–	–
GaitC3I-GB (ibid.)	68.9	80.4	–	–
GaitC3I (ibid.)	82.0	90.8	–	–
QAGait (Wang *et al.*, 2024)	59.1	74.0	–	–
VPNet (Ma *et al.*, 2024)	80.0	89.4	–	–
GaitSTR	**83.5**	**91.7**	**86.3**	**92.4**

6.2 GaitSTR Performance evaluation against the OU-ISIR Dataset

A comparative evaluation of various state-of-the-art gait recognition methods under different cross-view settings on the OU-ISIR large population dataset is presented in Table 3. Recognition accuracy (%) is reported for multiple gallery–probe angle combinations (55°, 65°, 75°, and 85°). Traditional approaches such as GEI (Yu *et al.*, 2006), SVD (Kusakunniran *et al.*, 2009), SVR (Kusakunniran *et al.*, 2010), and CMCC (ibid.) show significant performance drops under large view variations, while more advanced methods, including GEINet (Shiraga *et al.*, 2016), DCNN (Wu *et al.*, 2017) and DLWD (Wu *et al.*, 2018), achieve comparatively better results. Recent state-of-the-art techniques, such as TENFE (Singh & Goyal, 2020), GEI+MGANs (He *et al.*, 2019), and the proposed GaitSTR, maintain consistently high recognition rates across most angle pairs. Notably, GaitSTR attains the highest overall average accuracy of 95.2%, demonstrating superior robustness and adaptability to varying view angles compared to all other evaluated methods. The proposed method surpasses the second-best performer (TENFE, 93.9%) by a margin of 1.3%, establishing a new benchmark for cross-view gait recognition on this dataset.

6.3 GaitSTR Performance evaluation against the OU-MVLP Dataset

Table 4 presents a comparative evaluation of multiple state-of-the-art gait recognition methods on the OU-MVLP dataset across various view angles ranging from 0° to 270°. The recognition accuracies of earlier methods, such as GPAN (Chen *et al.*, 2022), GaitSet (Chao *et al.*, 2022), and MvGGAN (Chen *et al.*, 2021), demonstrate limited robustness under large view variations. More advanced techniques, including GaitGMT (Chen *et al.*, 2024), GaitAMR (Chen *et al.*, 2023), and GaitMPL (Dou *et al.*, 2024), achieve notable performance improvements through enhanced feature

Table 3. Accuracy (Average Rank-1) comparison on the OU-ISIR dataset. The table compares various methods under different gallery and probe angles.

Gallery angle(°)	55°			65°			75°			85°			Avg
Probe angle (°)	65°	75°	85°	55°	75°	85°	55°	65°	85°	55°	65°	75°	
TENFE (Singh & Goyal, 2020)	98.7	97.1	81.4	97.1	96.9	84.1	96.8	97.5	92.3	91.1	96.2	97.8	93.9
ffGEI (Wen & Wang, 2021)	59.3	57.8	63.1	60.2	64.8	66.3	61.5	65.7	68.3	62.9	63.7	67.4	63.3
DLWD (Wu *et al.*, 2018)	79.8	65.3	51.9	80.1	84.4	73.7	70.1	86.1	84.1	55.7	78.1	84.6	74.4
DCNN (Wu *et al.*, 2017)	98.3	96.0	80.5	96.3	97.3	83.3	94.2	97.8	92.4	90.0	96.0	98.4	93.3
GEINet (Shiraga *et al.*, 2016)	93.2	89.1	79.9	93.7	93.8	90.6	90.1	94.1	93.8	81.4	91.2	94.6	90.4
GEI+MGANs (He *et al.*, 2019)	99.0	96.1	77.9	97.7	98.5	84.4	94.8	98.9	86.4	86.9	97.4	99.5	93.1
CMCC (Kusakunniran *et al.*, 2010)	96.8	78.5	64.6	97.4	96.3	82.6	80.0	97.5	96.9	74.9	78.5	96.5	86.7
SVD (Kusakunniran *et al.*, 2009)	93.2	70.4	52.3	92.3	93.6	77.1	77.4	94.0	94.7	52.3	76.3	92.5	80.5
SVR (Kusakunniran *et al.*, 2010)	93.6	71.0	53.1	94.0	94.3	72.0	75.3	94.3	94.1	51.1	71.1	93.8	79.8
GEI (Yu *et al.*, 2006)	28.4	5.8	27.7	27.7	67.0	19.5	50.7	64.0	96.9	26.2	20.7	96.9	44.2
GaitSTR	**89.7**	**95.8**	**97.8**	**89.3**	**97.1**	**88.3**	**99.7**	**98.7**	**98.3**	**96.5**	**93.1**	**98.7**	**95.2**

representation and cross-view alignment. The highest overall performance is attained by the proposed **GaitSTR**, which achieves a mean accuracy of 96%, outperforming strong state-of-the-art baselines such as GEI-CNN (Elharrouss *et al.*, 2021) (95%) and Re-Id (Carley *et al.*, 2019) (92%). These results confirm GaitSTR's ability to advance the state-of-the-art in large-scale, cross-view gait recognition.

Table 4. Rank-1 Accuracy on the OU-MVLP dataset for various methods under different view angles.

Method	0°	15°	30°	45°	60°	75°	90°	180°	195°	210°	225°	240°	255°	270°	Avg
GTIEN (Chen & Li, 2024)	0.80	0.87	0.92	0.91	0.90	0.90	0.88	0.83	0.87	0.91	0.91	0.89	0.90	0.88	0.88
RDBA-Net (Junaid *et al.*, 2025)	0.83	0.89	0.90	0.91	0.89	0.90	0.89	0.85	0.88	0.90	0.90	0.89	0.88	0.88	0.89
GaitGMT (Chen *et al.*, 2024)	0.84	0.89	0.91	0.91	0.90	0.91	0.90	0.88	0.88	0.91	0.91	0.89	0.90	0.89	0.89
GaitAMR (Chen *et al.*, 2023)	0.84	0.89	0.89	0.90	0.88	0.89	0.88	0.86	0.88	0.88	0.88	0.87	0.89	0.87	0.88
PGOFI (Xu *et al.*, 2023)	0.79	0.86	0.88	0.89	0.86	0.87	0.86	0.84	0.85	0.87	0.89	0.85	0.85	0.86	0.86
GaitMPL(Dou *et al.*, 2024)	0.84	0.91	0.92	0.92	0.91	0.91	0.91	0.86	0.90	0.91	0.91	0.91	0.91	0.90	0.90
DANet (Ma *et al.*, 2023)	0.87	0.91	0.91	0.91	0.91	0.91	0.91	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.89
GPAN (Chen *et al.*, 2022)	0.69	0.81	0.87	0.87	0.81	0.85	0.82	0.73	0.79	0.85	0.85	0.80	0.83	0.80	0.81
GaitSet (Chao *et al.*, 2022)	0.79	0.87	0.89	0.90	0.88	0.88	0.87	0.81	0.86	0.89	0.89	0.87	0.87	0.86	0.87
MvGGAN (Chen *et al.*, 2021)	0.52	0.62	0.63	0.57	0.55	0.61	0.61	0.54	0.58	0.59	0.58	0.56	0.57	0.56	0.58
Re-Id (Carley *et al.*, 2019)	0.90	0.89	0.93	0.95	0.95	0.95	0.95	0.86	0.90	0.95	0.95	0.93	0.94	0.94	0.92
RPNet (Qin *et al.*, 2022)	0.73	0.84	0.89	0.89	0.86	0.87	0.86	0.76	0.83	0.88	0.88	0.85	0.86	0.84	0.85
GEI-CNN (Elharrouss *et al.*, 2021)	0.93	0.95	0.95	0.97	0.98	0.97	0.98	0.92	0.94	0.95	0.95	0.97	0.97	0.98	0.95
GaitSTR	**0.98**	**0.98**	**0.96**	**0.95**	**0.89**	**0.98**	**0.96**	**0.93**	**0.96**	**0.98**	**0.92**	**0.93**	**0.98**	**0.97**	**0.96**

6.4 GaitSTR Performance evaluation against the CASIA B Dataset

The comparison of state-of-the-art gait recognition methods on the CASIA-B dataset is presented in Table 5, across three modes—NM, BG and CL over eleven probe angles (0° to 180°).

Table 5. Rank-1 Accuracy on the CASIA-B dataset from different perspectives.

Mode	Method	0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°	Avg
NM	DensePoseGait (Liao *et al.*, 2025)	65.7	79.7	82.8	84.4	79.4	77.9	80.1	83.4	83.7	74.3	61.5	77.5
	ML-TAG (Saad Shakeel *et al.*, 2025)	96.4	97.8	98.9	97.8	97.4	97.0	97.7	99.0	98.8	99.1	95.3	97.0
	DDSTFDN (Qiao *et al.*, 2025)	94.2	98.1	98.4	98.0	96.9	96.3	99.5	99.3	98.7	99.1	94.2	97.2
	STTN (Chen & Li, 2024)	95.6	99.8	100.0	99.0	97.3	95.8	97.6	99.4	99.7	99.0	93.5	97.9
	LuGAN-HGC (Pan *et al.*, 2023)	89.3	88.1	89.0	89.9	87.4	88.7	87.4	88.8	88.8	87.0	87.0	88.3
	PGOFI (Xu *et al.*, 2023)	91.2	95.8	96.6	96.1	96.0	94.8	94.9	95.7	94.6	94.2	92.8	94.8
	GaitBase (Fan *et al.*, 2023)	94.8	99.7	99.8	99.0	96.8	95.3	97.3	99.2	99.6	99.2	94.8	97.8
	GaitSet (Chao *et al.*, 2022)	93.4	98.1	98.5	97.8	92.6	90.9	94.2	97.3	98.4	97.0	89.1	95.2
	GaitGraph (Teepe *et al.*, 2021)	85.3	88.5	91.0	92.5	87.2	86.5	88.4	89.2	87.9	85.9	81.9	87.7
	Siamese (Wang & Tang, 2021)	72.4	81.2	85.6	80.4	79.4	85.0	81.0	77.6	82.5	79.1	80.2	80.4
	GaitPart (Fan *et al.*, 2020)	94.1	98.6	99.3	98.5	94.0	92.3	95.9	98.4	99.2	97.8	90.4	96.2
	PoseGait (Liao *et al.*, 2020)	55.3	69.6	73.9	75.0	68.0	68.2	71.1	72.9	76.1	70.4	55.4	68.7
	GaitNet (Song *et al.*, 2019)	93.1	92.6	90.8	92.4	87.6	95.1	94.2	95.8	92.6	90.4	90.2	92.3
	CNN-LB (Wu *et al.*, 2017)	82.6	90.3	96.1	94.3	90.1	87.4	89.9	94.0	94.7	91.3	78.5	89.9
	GEINet (Shiraga *et al.*, 2016)	40.2	38.9	42.9	45.6	51.2	42.0	53.5	57.6	57.8	51.8	47.7	48.1
	GaitSTR	96.7	98.6	97.9	99.3	98.7	97.7	97.8	99.7	98.8	98.7	98.9	98.4
BG	DensePoseGait (Liao *et al.*, 2025)	55.4	70.4	76.6	73.3	65.6	65.3	68.1	71.0	69.8	57.3	44.8	65.2
	ML-TAG (Saad Shakeel *et al.*, 2025)	94.2	96.7	97.6	96.2	96.0	92.7	95.2	97.4	98.2	98.1	92.6	95.8
	DDSTFDN (Qiao *et al.*, 2025)	91.6	95.1	96.9	94.2	92.0	89.2	91.5	94.5	97.3	96.5	88.8	93.4
	STTN (Chen & Li, 2024)	92.4	95.7	97.0	96.0	92.5	89.6	91.7	96.7	98.8	98.0	88.5	94.3
	LuGAN-HGC (Pan *et al.*, 2023)	79.4	79.5	81.6	82.4	78.1	76.2	78.7	82.0	81.6	83.0	73.6	79.7
	PGOFI (Xu *et al.*, 2023)	87.6	90.8	91.7	91.5	91.0	93.9	90.1	91.5	92.0	90.4	89.5	90.9
	GaitBase (Fan *et al.*, 2023)	93.6	96.4	96.1	95.6	92.1	88.7	90.8	95.3	97.2	96.0	90.7	93.9
	GaitSet (Chao *et al.*, 2022)	85.9	92.1	93.9	90.4	86.4	78.7	85.0	91.6	93.1	91.0	80.7	88.1
	GaitGraph (Teepe *et al.*, 2021)	75.8	76.7	75.9	76.1	71.4	73.9	78.0	74.7	75.4	75.4	69.2	74.8
	Siamese (Wang & Tang, 2021)	62.5	68.7	69.4	64.8	62.8	67.2	68.3	65.7	60.7	64.1	60.3	65.0
	GaitPart (Fan *et al.*, 2020)	89.1	94.8	96.7	95.1	88.3	84.9	89.0	93.5	96.1	93.8	85.8	91.6
	PoseGait (Liao *et al.*, 2020)	35.3	47.2	52.4	46.9	45.5	43.9	46.1	48.1	49.4	43.6	31.1	44.5
	GaitNet (Song *et al.*, 2019)	88.8	88.7	88.7	94.3	85.4	92.7	91.1	92.6	84.9	84.4	86.7	88.9
	CNN-LB (Wu *et al.*, 2017)	64.2	80.6	82.7	76.9	64.8	63.1	68.0	76.9	82.2	75.4	61.3	72.4
	GEINet (Shiraga *et al.*, 2016)	34.2	29.3	31.2	35.2	35.2	27.6	35.9	43.5	45.0	39.0	36.8	35.7
	GaitSTR	97.6	96.7	95.6	97.8	93.2	91.8	98.7	96.7	95.4	94.8	96.8	95.9
CL	DensePoseGait (Liao *et al.*, 2025)	41.8	47.7	49.7	50.3	46.5	46.0	49.5	47.8	47.4	39.4	29.3	45.2
	ML-TAG (Saad Shakeel *et al.*, 2025)	76.6	91.1	93.3	90.0	86.7	81.0	85.4	89.3	90.3	87.4	72.2	85.7
	DDSTFDN (Qiao *et al.*, 2025)	70.1	83.4	84.6	81.2	79.2	74.2	76.0	81.0	83.9	80.6	67.0	78.3
	STTN (Chen & Li, 2024)	69.7	89.0	88.4	84.9	78.8	75.5	79.2	82.4	82.6	76.9	61.9	79.0
	LuGAN-HGC (Pan *et al.*, 2023)	72.8	72.3	69.4	75.2	77.0	79.6	80.5	78.1	76.3	74.9	72.8	75.4
	PGOFI (Xu *et al.*, 2023)	73.0	74.5	79.1	79.8	81.5	82.5	81.1	79.4	77.8	76.6	75.7	78.3
	GaitBase (Fan *et al.*, 2023)	68.8	81.7	84.8	81.7	79.0	75.7	78.0	80.7	82.2	78.3	66.8	78.0
	GaitSet (Chao *et al.*, 2022)	63.7	75.6	80.7	77.5	69.1	67.8	69.7	74.6	76.1	71.1	55.7	71.1
	GaitGraph (Teepe *et al.*, 2021)	69.6	66.1	68.8	67.2	64.5	62.0	69.5	65.6	65.7	66.1	64.3	66.3
	Siamese (Wang & Tang, 2021)	57.8	63.2	68.3	64.1	66.0	64.8	67.7	60.2	66.0	68.3	60.3	64.2
	GaitPart (Fan *et al.*, 2020)	70.7	85.5	86.9	83.3	77.1	72.5	76.9	82.2	83.8	80.2	66.5	78.7
	PoseGait (Liao *et al.*, 2020)	24.3	29.7	41.3	38.8	38.2	38.5	41.6	44.9	42.2	33.4	22.5	36.0
	GaitNet (Song *et al.*, 2019)	50.1	60.7	72.4	72.1	74.6	78.4	70.3	68.2	53.5	44.1	40.8	62.3
	CNN-LB (Wu *et al.*, 2017)	37.7	57.2	66.6	61.1	55.2	54.6	55.2	59.1	58.9	48.8	39.4	54.0
	GEINet (Shiraga *et al.*, 2016)	19.9	20.3	22.5	23.5	26.7	21.3	27.4	28.2	24.2	22.5	21.6	23.5
	GaitSTR	87.1	89.8	92.5	93.3	86.2	77.3	77.9	85.3	93.9	81.3	69.7	84.9

In the NM mode, **GaitSTR** achieves the highest overall performance with an average accuracy of 98.1%, closely followed by (Chen & Li, 2024) at 97.9% and GaitPart (Fan *et al.,* 2020) at 96.2%. Traditional appearance-based models such as GEINet (Shiraga *et al.,* 2016) exhibit significantly lower performance, with an average of 48.1%, indicating the superiority of modern deep spatio-temporal feature learning strategies.

For the BG mode, **GaitSTR** again outperforms other methods, achieving 95.3% average accuracy, followed by STTN(94.3%) and GaitPart (91.6%). Although PGOFI (Xu *et al.,* 2023) maintains competitive results (90.9%), appearance-dependent methods like PoseGait (Liao *et al.,* 2020) and GEINet (Shiraga *et al.,* 2016) record much lower averages (36.0% and 23.5%, respectively), highlighting their vulnerability to occlusions caused by carried objects.

Under the challenging CL mode, where gait silhouettes are heavily occluded by clothing variations, **GaitSTR** maintains robust recognition with an average of 83.7%, outperforming all baselines. STTN records 79.0%, and PGOFI achieves 78.3%, while traditional CNN-based approaches such as CNN-LB (Wu *et al.,* 2017) drop to 54.0% and GEINet remains below 25%, confirming the difficulty of handling large clothing variations.

6.5 Ablation Study

In this subsection, we conduct ablation studies to evaluate the individual contributions of the channel-space attention module, temporal attention module, and the effect of data augmentation in our proposed framework. All ablation experiments are conducted on the CASIA-B dataset for the three different modes.

Ablation study on different modules. To explore the contribution of each attention component, we start with a baseline network without any attention or augmentation, then gradually add channel-space attention (CSA), temporal attention (TA), and data augmentation (DA). The ablation results are summarized in Table 6. **1) Effectiveness of Channel-Space Attention.** As shown in Table 6, adding CSA to the baseline yields noticeable gains in Rank-1 accuracy across all conditions, improving the mean accuracy from 82.8% to 88.1%. This demonstrates that enhancing channel and spatial dependencies helps the network focus on discriminative gait regions. **(2) Effectiveness of Temporal Attention.** Integrating TA into the baseline improves recognition by leveraging temporal dependencies between frames, achieving a mean accuracy of 87.7%. This confirms that certain gait frames are more informative for identity recognition. **(3) Combined Effect of CSA and TA.** When both CSA and TA are used, the mean accuracy further improves to 89.8%, indicating that spatial-channel refinement and temporal weighting are complementary. **(4) Effectiveness of Data Augmentation.** Finally, applying data augmentation to the full model boosts the mean accuracy to 93.0%, highlighting its importance in improving robustness against variations such as occlusion, speed changes, and clothing. These results clearly demonstrate that both CSA and TA independently enhance recognition performance, while their combination produces further gains. Moreover, applying data augmentation significantly strengthens robustness across all conditions.

Table 6. Ablation study of channel-space attention (CSA), temporal attention (TA), and data augmentation (DA) on CASIA-B dataset (Rank-1, %).

Structure	CSA	TA	DA	NM (%)	BG (%)	CL (%)	Mean (%)
Baseline	–	–	–	88.1	86.2	74.2	82.8
CSA only	✓	–	–	92.3	90.4	81.7	88.1
TA only	–	✓	–	92.7	91.0	79.5	87.7
CSA + TA	✓	✓	–	95.1	92.5	82.0	89.8
Without Data Augmentation	–	✓	✓	96.2	93.3	82.9	90.8
Full (CSA+TA+DA)	✓	✓	✓	98.4	95.9	84.9	93.0

7. Conclusions

In this paper, we proposed GaitSTR, a novel gait recognition framework that integrates pyramid-based hierarchical feature extraction with attention-guided spatial and temporal modeling. GaitSTR decomposes gait sequences into multi-scale spatial representations, capturing both fine- and coarse-grained motion patterns, while a memory-augmented RNN with temporal attention effectively models sequential dynamics. The attention-guided dense network enhances spatial feature learning by focusing on the most informative regions within silhouettes. Extensive experiments conducted on four widely used benchmark datasets—GREW, OU-ISIR, OU-MVLP, and CASIA-B—demonstrated that GaitSTR consistently outperforms state-of-the-art methods, achieving notable improvements under variations in clothing, carrying conditions, and viewpoints. These results validate the robustness and generalization capability of the proposed approach, establishing GaitSTR as a strong benchmark for future gait recognition research. Future research will aim to extend GaitSTR's capability to handle unseen environments and extreme covariates through advanced cross-domain learning techniques.

Acknowledgments

The authors express their sincere gratitude to the Department of Electronics and Telecommunication Engineering, Sanjivani College of Engineering, Kopargaon, for their constant support and encouragement. They also extend their thanks to the Co-Editor-in-Chief, Prof. João Domingos Scalon, and the esteemed reviewers for their valuable and constructive feedback, which has significantly contributed to improving the clarity, quality, and presentation of this manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

Conceptualization: MANDLIK, S.B.; LABADE, R.P. **Data curation**: MANDLIK, S.B.; LABADE, R.P. **Formal analysis**: MANDLIK, S.B.; LABADE, R.P.; CHAUDHARI, S.V. **Investigation**: MANDLIK, S.B.; LABADE, R.P. **Methodology**: MANDLIK, S.B.; LABADE, R.P. **Project administration**: MANDLIK, S.B.; LABADE, R.P.; CHAUDHARI, S.V. **Software**: MANDLIK, S.B.; LABADE, R.P. **Resources**: MANDLIK, S.B.; LABADE, R.P.; CHAUDHARI, S.V. **Validation**: MANDLIK, S.B.; LABADE, R.P. **Visualization**: MANDLIK, S.B.; LABADE, R.P. **Writing - original draft**: MANDLIK, S.B.; LABADE, R.P. **Writing - review and editing**: MANDLIK, S.B.; LABADE, R.P.; CHAUDHARI, S.V.; AGARKAR, B.S.

References

1. Altab Hossain, M., Makihara, Y., Wang, J. & Yagi, Y. Clothing-invariant gait identification using part-based clothing categorization and adaptive weight control. *Pattern Recognition* **43,** 2281–2291. https://doi.org/10.1016/J.PATCOG.2009.12.020 (June 2010).
2. Balazia, M. & Sojka, P. Gait Recognition from Motion Capture Data. *ACM Transactions on Multimedia Computing, Communications, and Applications* **14,** 1–18. https://doi.org/10.1145/3152124 (Mar. 2018).
3. Bastos, D. R. M. & Tavares, J. M. R. S. A scalable gait acquisition and recognition system with angle-enhanced models. *Expert Systems with Applications* **269,** 126499. https://doi.org/10.1016/J.ESWA.2025.126499 (Apr. 2025).

4. Ben, X., Gong, C., Zhang, P., Yan, R., Wu, Q. & Meng, W. Coupled Bilinear Discriminant Projection for Cross-View Gait Recognition. *IEEE Transactions on Circuits and Systems for Video Technology* **30,** 734–747. https://doi.org/10.1109/TCSVT.2019.2893736 (Mar. 2020).
5. BenAbdelkader, C., Cutler, R. & Davis, L. *Stride and cadence as a biometric in automatic person identification and verification* in *Proceedings of Fifth IEEE International Conference on Automatic Face and Gesture Recognition* (IEEE, 2002), 372–377. https://doi.org/10.1109/AFGR.2002.1004182.
6. Bouchrika, I. & Nixon, M. S. in *Computer Vision/Computer Graphics Collaboration Techniques* 150–160 (Springer Berlin Heidelberg, 2008). https://doi.org/10.1007/978-3-540-71457-6_14.
7. Cai, S., Chen, D., Fan, B., Du, M., Bao, G. & Li, G. Gait phases recognition based on lower limb sEMG signals using LDA-PSO-LSTM algorithm. *Biomedical Signal Processing and Control* **80,** 104272. https://doi.org/10.1016/J.BSPC.2022.104272 (Feb. 2023).
8. Carley, C., Ristani, E. & Tomasi, C. *Person Re-Identification From Gait Using an Autocorrelation Network* in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (IEEE, June 2019), 2345–2353. https://doi.org/10.1109/CVPRW.2019.00288.
9. Castro, F. M., Delgado-Escaño, R., Hernández-García, R., Marín-Jiménez, M. J. & Guil, N. AttenGait: Gait recognition with attention and rich modalities. *Pattern Recognition* **148,** 110171. https://doi.org/10.1016/j.patcog.2023.110171 (Apr. 2024).
10. Chao, H., Wang, K., He, Y., Zhang, J. & Feng, J. GaitSet: Cross-View Gait Recognition Through Utilizing Gait As a Deep Set. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **44,** 3467–3478. https://doi.org/10.1109/TPAMI.2021.3057879 (2022).
11. Chen, G. *et al.* GaitGMT: Global feature mapping transformer for gait recognition. *Journal of Visual Communication and Image Representation* **100,** 104139. https://doi.org/10.1016/j.jvcir.2024.104139 (Apr. 2024).
12. Chen, J., Wang, Z., Yi, P., Zeng, K., He, Z. & Zou, Q. Gait Pyramid Attention Network: Toward Silhouette Semantic Relation Learning for Gait Recognition. *IEEE Transactions on Biometrics, Behavior, and Identity Science* **4,** 582–595. https://doi.org/10.1109/TBIOM.2022.3213545 (Oct. 2022).
13. Chen, J., Wang, Z., Zheng, C., Zeng, K., Zou, Q. & Cui, L. GaitAMR: Cross-view gait recognition via aggregated multi-feature representation. *Information Sciences* **636,** 118920. https://doi.org/10.1016/J.INS.2023.03.145 (July 2023).
14. Chen, X., Luo, X., Weng, J., Luo, W., Li, H. & Tian, Q. Multi-View Gait Image Generation for Cross-View Gait Recognition. *IEEE Transactions on Image Processing* **30,** 3041–3055. https://doi.org/10.1109/TIP.2021.3055936 (2021).
15. Chen, X., Weng, J., Lu, W. & Xu, J. Multi-Gait Recognition Based on Attribute Discovery. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **40,** 1697–1710. https://doi.org/10.1109/TPAMI.2017.2726061 (2018).
16. Chen, Y. & Li, X. Gait feature learning via spatio-temporal two-branch networks. *Pattern Recognition* **147,** 110090. https://doi.org/10.1016/j.patcog.2023.110090 (Mar. 2024).
17. Choi, S., Kim, J., Kim, W. & Kim, C. Skeleton-Based Gait Recognition via Robust Frame-Level Matching. *IEEE Transactions on Information Forensics and Security* **14,** 2577–2592. https://doi.org/10.1109/TIFS.2019.2901823 (Oct. 2019).
18. Cunado, D., Nixon, M. S. & Carter, J. N. Automatic extraction and description of human gait models for recognition purposes. *Computer Vision and Image Understanding* **90,** 1–41. https://doi.org/10.1016/S1077-3142(03)00008-0 (Apr. 2003).

19. Deng, J., Guo, J., Xue, N. & Zafeiriou, S. *ArcFace: Additive Angular Margin Loss for Deep Face Recognition* in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, June 2019), 4685–4694. https://doi.org/10.1109/CVPR.2019.00482.

20. Dou, H., Zhang, P., Su, W., Yu, Y., Lin, Y. & Li, X. *GaitGCI: Generative Counterfactual Intervention for Gait Recognition* in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, June 2023), 5578–5588. https://doi.org/10.1109/CVPR52729.2023.00540.

21. Dou, H., Zhang, P., Zhao, Y., Dong, L., Qin, Z. & Li, X. GaitMPL: Gait Recognition With Memory-Augmented Progressive Learning. *IEEE Transactions on Image Processing* **33,** 1464–1475. https://doi.org/10.1109/TIP.2022.3164543 (2024).

22. 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* **34,** 4230–4241. https://doi.org/10.1109/TIP.2024.3360870 (2025).

23. Du, Z. & Zhao, D. *Deep Learning Gait Recognition Method with Self-attention* in *2024 4th International Conference on Artificial Intelligence, Robotics, and Communication (ICAIRC)* (IEEE, Dec. 2024), 933–937. https://doi.org/10.1109/ICAIRC64177.2024.10900018.

24. Elharrouss, O., Almaadeed, N., Al-Maadeed, S. & Bouridane, A. Gait recognition for person re-identification. *Journal of Supercomputing* **77,** 3653–3672. https://doi.org/10.1007/s11227-020-03409-5 (2021).

25. Erdaş, B., Sümer, E. & Kibaroğlu, S. Neurodegenerative disease detection and severity prediction using deep learning approaches. *Biomedical Signal Processing and Control* **70,** 103069. https://doi.org/10.1016/J.BSPC.2021.103069 (Sept. 2021).

26. Fan, C., Liang, J., Shen, C., Hou, S., Huang, Y. & Yu, S. *OpenGait: Revisiting Gait Recognition Toward Better Practicality* in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, June 2023), 9707–9716. https://doi.org/10.1109/CVPR52729.2023.00936.

27. Fan, C. *et al. GaitPart: Temporal part-based model for gait recognition* in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, June 2020), 14213–14221. https://doi.org/10.1109/CVPR42600.2020.01423.

28. Filipi Gonçalves Dos Santos, C. *et al.* Gait Recognition Based on Deep Learning: A Survey. https://doi.org/10.1145/3490235 (Feb. 2023).

29. Gadaleta, M. & Rossi, M. IDNet: Smartphone-based gait recognition with convolutional neural networks. *Pattern Recognition* **74,** 25–37. https://doi.org/10.1016/J.PATCOG.2017.09.005 (Feb. 2018).

30. Gao, S., Yun, J., Zhao, Y. & Liu, L. Gait-D: Skeleton-based gait feature decomposition for gait recognition. *IET Computer Vision* **16,** 111–125. https://doi.org/10.1049/cvi2.12070 (Mar. 2022).

31. Guo, X. *et al.* Gait Recognition in the Wild: A Large-scale Benchmark and NAS-based Baseline. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 1–18. https://doi.org/10.1109/TPAMI.2025.3546482 (2025).

32. Gupta, S. K. & Chattopadhyay, P. Gait recognition in the presence of co-variate conditions. *Neurocomputing* **454,** 76–87. https://doi.org/10.1016/J.NEUCOM.2021.04.113 (Sept. 2021).

33. Han, J. J. & Bhanu, B. Individual recognition using gait energy image. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28,** 316–322. https://doi.org/10.1109/TPAMI.2006.38 (Feb. 2006).

34. Hasan, K., Uddin, M. Z., Ray, A., Hasan, M., Alnajjar, F. & Ahad, M. A. R. Improving Gait Recognition Through Occlusion Detection and Silhouette Sequence Reconstruction. *IEEE Access* **12,** 158597–158610. https://doi.org/10.1109/ACCESS.2024.3482430 (2024).

35. He, Y., Zhang, J., Shan, H. & Wang, L. Multi-Task GANs for view-specific feature learning in gait recognition. *IEEE Transactions on Information Forensics and Security* **14,** 102–113. https://doi.org/10.1109/TIFS.2018.2844819 (2019).

36. Hou, S., Liu, X., Cao, C. & Huang, Y. Gait Quality Aware Network: Toward the Interpretability of Silhouette-Based Gait Recognition. *IEEE Transactions on Neural Networks and Learning Systems* **34,** 8978–8988. https://doi.org/10.1109/TNNLS.2022.3154723 (Nov. 2023).

37. Huang, P. *et al.* in *Advances in Biometrics* 380–397 (Springer, 2025). https://doi.org/10.1007/978-3-031-72658-3_22.

38. Huang, T., Ben, X., Gong, C., Xu, W., Wu, Q. & Zhou, H. GaitDAN: Cross-View Gait Recognition via Adversarial Domain Adaptation. *IEEE Transactions on Circuits and Systems for Video Technology* **34,** 8026–8040. https://doi.org/10.1109/TCSVT.2024.3384308 (Sept. 2024).

39. Huang, X. *et al. Context-Sensitive Temporal Feature Learning for Gait Recognition* in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)* (IEEE, Oct. 2021), 12889–12898. https://doi.org/10.1109/ICCV48922.2021.01267.

40. Huang, X. & Boulgouris, N. V. Gait Recognition With Shifted Energy Image and Structural Feature Extraction. *IEEE Transactions on Image Processing* **21,** 2256–2268. https://doi.org/10.1109/TIP.2011.2180914 (Apr. 2012).

41. Huo, W., Tang, J., Bao, W., Wang, K., Wang, N. & Liang, D. Multiple motion pattern augmentation assisted gait recognition. *Signal Processing* **238,** 110185. https://doi.org/10.1016/j.sigpro.2025.110185 (Jan. 2026).

42. Iwama, H., Okumura, M., Makihara, Y. & Yagi, Y. The OU-ISIR Gait Database Comprising the Large Population Dataset and Performance Evaluation of Gait Recognition. *IEEE Transactions on Information Forensics and Security* **7,** 1511–1521. https://doi.org/10.1109/TIFS.2012.2204253 (Oct. 2012).

43. Johansson, G. Visual perception of biological motion and a model for its analysis. *Perception Psychophysics* **14,** 201–211. https://doi.org/10.3758/BF03212378 (June 1973).

44. Junaid *et al.* Human gait recognition using dense residual network and hybrid attention technique with back-flow mechanism. *Digital Signal Processing* **166,** 105401. https://doi.org/10.1016/j.dsp.2025.105401 (Nov. 2025).

45. Kusakunniran, W., Wu, Q., Li, H. & Zhang, J. *Multiple views gait recognition using View Transformation Model based on optimized Gait Energy Image* in *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops* (IEEE, Sept. 2009), 1058–1064. https://doi.org/10.1109/ICCVW.2009.5457587.

46. Kusakunniran, W., Wu, Q., Zhang, J. & Li, H. *Support vector regression for multi-view gait recognition based on local motion feature selection* in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (IEEE, June 2010), 974–981. https://doi.org/10.1109/CVPR.2010.5540113.

47. Li, J., Wang, Z., Wang, C. & Su, W. GaitFormer: Leveraging dual-stream spatial–temporal Vision Transformer via a single low-cost RGB camera for clinical gait analysis. *Knowledge-Based Systems* **295,** 111810. https://doi.org/10.1016/J.KNOSYS.2024.111810 (July 2024).

48. Li, T. *et al.* A survey on gait recognition against occlusion: taxonomy, dataset and methodology. *PeerJ Computer Science* **10,** e2602. https://doi.org/10.7717/peerj-cs.2602 (Dec. 2024).

49. Li, X., Makihara, Y., Xu, C., Yagi, Y. & Ren, M. Joint Intensity Transformer Network for Gait Recognition Robust Against Clothing and Carrying Status. *IEEE Transactions on Information Forensics and Security* **14,** 3102–3115. https://doi.org/10.1109/TIFS.2019.2912577 (Dec. 2019).

50. Liao, R., Li, Z., Bhattacharyya, S. S. & York, G. DensePoseGait: Dense Human Pose Part-Guided for Gait Recognition. *IEEE Transactions on Biometrics, Behavior, and Identity Science* **7,** 33–46. https://doi.org/10.1109/TBIOM.2024.3486732 (Jan. 2025).

51. Liao, R., Yu, S., An, W. & Huang, Y. A model-based gait recognition method with body pose and human prior knowledge. *Pattern Recognition* **98,** 107069. https://doi.org/10.1016/j.patcog.2019.107069 (Feb. 2020).

52. Lin, B., Zhang, S. & Yu, X. *Gait Recognition via Effective Global-Local Feature Representation and Local Temporal Aggregation* in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (IEEE, 2021), 14628–14636. https://doi.org/10.1109/ICCV48922.2021.01438.

53. Liu, Y., Jiang, X., Sun, T. & Xu, K. *3D Gait Recognition Based on a CNN-LSTM Network with the Fusion of SkeGEI and DA Features* in *2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)* (IEEE, Sept. 2019), 1–8. https://doi.org/10.1109/AVSS.2019.8909881.

54. Liu, Y., Zhang, Y., Bhanu, B., Coleman, S. & Kerr, D. Multi-level cross-view consistent feature learning for person re-identification. *Neurocomputing* **435,** 1–14. https://doi.org/10.1016/J.NEUCOM.2021.01.010 (May 2021).

55. Liu, Z. & Sarkar, S. Improved gait recognition by gait dynamics normalization. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28,** 863–876. https://doi.org/10.1109/TPAMI.2006.122 (2006).

56. Ma, G., Wu, L. & Wang, Y. A general subspace ensemble learning framework via totally-corrective boosting and tensor-based and local patch-based extensions for gait recognition. *Pattern Recognition* **66,** 280–294. https://doi.org/10.1016/J.PATCOG.2017.01.003 (June 2017).

57. Ma, K., Fu, Y., Cao, C., Hou, S., Huang, Y. & Zheng, D. *Learning Visual Prompt for Gait Recognition* in *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, June 2024), 593–603. https://doi.org/10.1109/CVPR52733.2024.00063.

58. Ma, K., Fu, Y., Zheng, D., Cao, C., Hu, X. & Huang, Y. *Dynamic Aggregated Network for Gait Recognition* in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, June 2023), 22076–22085. https://doi.org/10.1109/CVPR52729.2023.02114.

59. Makihara, Y., Mansur, A., Muramatsu, D., Uddin, Z. & Yagi, Y. *Multi-view discriminant analysis with tensor representation and its application to cross-view gait recognition* in *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)* (IEEE, May 2015), 1–8. https://doi.org/10.1109/FG.2015.7163131.

60. Mandlik, S., Labade, R., Chaudhari, S. & Agarkar, B. Enhancing Gait Recognition with Attention-Based Spatial-Temporal Deep Learning: The GaitDeep Framework. *Computer Science Journal of Moldova* **33,** 188–218. https://doi.org/10.56415/csjm.v33.10 (2025).

61. Mandlik, S., Labade, R., Chaudhari, S. & Agarkar, B. *GRDDN: Enhanced Gait Recognition using a Deep Dense Network* in *2025 International Conference on Inventive Computation Technologies (ICICT)* (IEEE, Apr. 2025), 129–135. https://doi.org/10.1109/ICICT64420.2025.11005378.

62. Mandlik, S. B., Labade, R., Chaudhari, S. V. & Agarkar, B. S. Review of gait recognition systems: approaches and challenges. *International Journal of Electrical and Computer Engineering* **15,** 349. https://doi.org/10.11591/ijece.v15i1.pp349-355 (Feb. 2025).

63. Mitra, S. & Acharya, T. Gesture Recognition: A Survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* **37,** 311–324. https://doi.org/10.1109/TSMCC.2007.893280 (May 2007).

64. Mizuno, M., Fujita, T., Kawanishi, Y., Deguchi, D. & Murase, H. Subjective Baggage-Weight Estimation Based on Human Walking Behavior. *IEEE Access* **12,** 39390–39398. https://doi.org/10.1109/ACCESS.2024.3376656 (2024).

65. Mogan, J. N., Lee, C. P. & Lim, K. M. Ensemble CNN-ViT Using Feature-Level Fusion for Gait Recognition. *IEEE Access* **12,** 108573–108583. https://doi.org/10.1109/ACCESS.2024.3439602 (2024).

66. Muramatsu, D., Shiraishi, A., Makihara, Y., Uddin, M. Z. & Yagi, Y. Gait-Based Person Recognition Using Arbitrary View Transformation Model. *IEEE Transactions on Image Processing* **24,** 140–154. https://doi.org/10.1109/TIP.2014.2371335 (Jan. 2015).

67. Niyogi, S. & Adelson, E. *Analyzing and recognizing walking figures in XYT* in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE Computer Society Press, 1994), 469–474. https://doi.org/10.1109/CVPR.1994.323868.

68. Pan, H., Chen, Y., Xu, T., He, Y. & He, Z. Toward Complete-View and High-Level Pose-Based Gait Recognition. *IEEE Transactions on Information Forensics and Security* **18,** 2104–2118. https://doi.org/10.1109/TIFS.2023.3254449 (2023).

69. Panahi, L. & Ghods, V. Human fall detection using machine vision techniques on RGB–D images. *Biomedical Signal Processing and Control* **44,** 146–153. https://doi.org/10.1016/J.BSPC.2018.04.014 (July 2018).

70. Pinčić, D., Sušanj, D. & Lenac, K. Gait Recognition with Self-Supervised Learning of Gait Features Based on Vision Transformers. *Sensors* **22,** 7140. https://doi.org/10.3390/s22197140 (Sept. 2022).

71. Prajapati, N., Kaur, A. & Sethi, D. *A Review on Clinical Gait Analysis* in *2021 5th International Conference on Trends in Electronics and Informatics (ICOEI)* (IEEE, June 2021), 967–974. https://doi.org/10.1109/ICOEI51242.2021.9452951.

72. Qiao, S., Tang, C., Hu, H., Wang, W., Tong, A. & Ren, F. Cross-view identification based on gait bioinformation using a dynamic densely connected spatial-temporal feature decoupling network. *Biomedical Signal Processing and Control* **104,** 107494. https://doi.org/10.1016/j.bspc.2025.107494 (June 2025).

73. Qin, H., Chen, Z., Guo, Q., Wu, Q. M. J. & Lu, M. RPNet: Gait Recognition With Relationships Between Each Body-Parts. *IEEE Transactions on Circuits and Systems for Video Technology* **32,** 2990–3000. https://doi.org/10.1109/TCSVT.2021.3095290 (May 2022).

74. Rashmi, M. & Guddeti, R. M. R. Human identification system using 3D skeleton-based gait features and LSTM model. *Journal of Visual Communication and Image Representation* **82,** 103416. https://doi.org/10.1016/J.JVCIR.2021.103416 (Jan. 2022).

75. Saad Shakeel, M., Liu, K., Liao, X. & Kang, W. TAG: A Temporal Attentive Gait Network for Cross-View Gait Recognition. *IEEE Transactions on Instrumentation and Measurement* **74,** 1–14. https://doi.org/10.1109/TIM.2024.3497164 (2025).

76. Sepas-Moghaddam, A. & Etemad, A. Deep Gait Recognition: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **45,** 264–284. https://doi.org/10.1109/TPAMI.2022.3151865 (Jan. 2023).

77. Sethi, D., Prakash, C. & Bharti, S. in *Proceedings of the 2022 International Conference on Artificial Intelligence and Computer Vision (AICV 2022)* 363–375 (Springer, 2022). https://doi.org/10.1007/978-3-030-95711-7_31.

78. Shiraga, K., Makihara, Y., Muramatsu, D., Echigo, T. & Yagi, Y. *GEINet: View-invariant gait recognition using a convolutional neural network* in *2016 International Conference on Biometrics (ICB)* (IEEE, June 2016), 1–8. https://doi.org/10.1109/ICB.2016.7550060.

79. Singh, J. & Goyal, G. Identifying biometrics in the wild – A time, erosion and neural inspired framework for gait identification. *Journal of Visual Communication and Image Representation* **66,** 102725. https://doi.org/10.1016/j.jvcir.2019.102725 (Jan. 2020).

80. Sokolova, A. & Konushin, A. Methods of Gait Recognition in Video. *Programmirovanie* **45,** 213–220. https://doi.org/10.1134/S0361768819040091 (2019).

81. Song, C., Huang, Y., Huang, Y., Jia, N. & Wang, L. GaitNet: An end-to-end network for gait based human identification. *Pattern Recognition* **96,** 106988. https://doi.org/10.1016/J.PATCOG.2019.106988 (Dec. 2019).

82. Song, X. *et al.* Gait Attribute Recognition: A New Benchmark for Learning Richer Attributes From Human Gait Patterns. *IEEE Transactions on Information Forensics and Security* **19,** 1–14. https://doi.org/10.1109/TIFS.2023.3318934 (2024).

83. Takemura, N., Makihara, Y., Muramatsu, D., Echigo, T. & Yagi, Y. Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition. *IPSJ Transactions on Computer Vision and Applications* **10.** https://doi.org/10.1186/s41074-018-0039-6 (2018).

84. Teepe, T. *et al. GaitGraph: Graph Convolutional Network for Skeleton-Based Gait Recognition* in *2021 IEEE International Conference on Image Processing (ICIP)* (IEEE, Jan. 2021), 2314–2318. https://doi.org/10.1109/ICIP42928.2021.9506717.

85. Tong, S., Fu, Y. & Ling, H. *Verification-based pairwise gait identification* in *2017 IEEE International Conference on Multimedia Expo Workshops (ICMEW)* (IEEE, July 2017), 669–673. https://doi.org/10.1109/ICMEW.2017.8026299.

86. Uddin, M. Z. *et al.* The OU-ISIR Large Population Gait Database with real-life carried object and its performance evaluation. *IPSJ Transactions on Computer Vision and Applications* **10,** 5. https://doi.org/10.1186/s41074-018-0041-z (Dec. 2018).

87. Wang, J. *et al. Causal Intervention for Sparse-View Gait Recognition* in *Proceedings of the 31st ACM International Conference on Multimedia* (ACM, Oct. 2023), 77–85. https://doi.org/10.1145/3581783.3612124.

88. Wang, J. *et al.* GaitC 3 I: Robust Cross-Covariate Gait Recognition via Causal Intervention. *IEEE Transactions on Circuits and Systems for Video Technology,* 1–1. https://doi.org/10.1109/TCSVT.2025.3545210 (2025).

89. Wang, L., Liu, B., Liang, F. & Wang, B. *Hierarchical Spatio-Temporal Representation Learning for Gait Recognition* in *2023 IEEE/CVF International Conference on Computer Vision (ICCV)* (IEEE, Oct. 2023), 19582–19592. https://doi.org/10.1109/ICCV51070.2023.01799.

90. Wang, M. *et al. DyGait: Exploiting Dynamic Representations for High-performance Gait Recognition* in *2023 IEEE/CVF International Conference on Computer Vision (ICCV)* (IEEE, Oct. 2023), 13378–13387. https://doi.org/10.1109/ICCV51070.2023.01235.

91. Wang, Y., Chen, Z., Wu, Q. M. J. & Rong, X. Deep mutual learning network for gait recognition. *Multimedia Tools and Applications* **79,** 22653–22672. https://doi.org/10.1007/s11042-020-09003-4 (2020).

92. Wang, Y., Liu, B., Zhao, Z., Niu, J., Chu, Q. & Yu, N. *CMGait: Enhancing Cross-Modality Gait Recognition between LiDAR and RGB through Contrastive Identity-consistent Feature Aggregation* in *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (IEEE, Apr. 2025), 1–5. https://doi.org/10.1109/ICASSP49660.2025.10889323.

93. Wang, Z. & Tang, C. *Model-based gait recognition using graph network on very large population database* Dec. 2021.

94. Wang, Z. *et al. QAGait: Revisit Gait Recognition from a Quality Perspective* in *Proceedings of the AAAI Conference on Artificial Intelligence* **38** (AAAI, Mar. 2024), 5785–5793. https://doi.org/10.1609/aaai.v38i6.28391.

95. Wei, T., Liu, M., Zhao, H. & Li, H. GMSN: An efficient multi-scale feature extraction network for gait recognition. *Expert Systems with Applications* **252,** 124250. https://doi.org/10.1016/j.eswa.2024.124250 (Oct. 2024).

96. Wen, J. & Wang, X. Gait recognition based on sparse linear subspace. *IET Image Processing* **15,** 2761–2769. https://doi.org/10.1049/ipr2.12260 (Oct. 2021).

97. Wu, H., Weng, J., Chen, X. & Lu, W. Feedback weight convolutional neural network for gait recognition. *Journal of Visual Communication and Image Representation* **55,** 424–432. https://doi.org/10.1016/J.JVCIR.2018.06.019 (Aug. 2018).

98. Wu, Z., Huang, Y., Wang, L., Wang, X. & Tan, T. A Comprehensive Study on Cross-View Gait Based Human Identification with Deep CNNs. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **39,** 209–226. https://doi.org/10.1109/TPAMI.2016.2545669 (Feb. 2017).

99. Xing, W., Li, Y. & Zhang, S. View-invariant gait recognition method by three-dimensional convolutional neural network. *Journal of Electronic Imaging* **27,** 1. https://doi.org/10.1117/1.JEI.27.1.013010 (Jan. 2018).

100. Xiong, H., Feng, B., Wang, X. & Liu, W. in *Advances in Biometrics* 251–270 (Springer, 2025). https://doi.org/10.1007/978-3-031-72949-2_15.

101. Xu, C., Makihara, Y., Li, X., Yagi, Y. & Lu, J. Cross-View Gait Recognition Using Pairwise Spatial Transformer Networks. *IEEE Transactions on Circuits and Systems for Video Technology* **31,** 260–274. https://doi.org/10.1109/TCSVT.2020.2975671 (Jan. 2021).

102. Xu, J., Li, H. & Hou, S. Attention-based gait recognition network with novel partial representation PGOFI based on prior motion information. *Digital Signal Processing* **133,** 103845. https://doi.org/10.1016/j.dsp.2022.103845 (Mar. 2023).

103. Yao, L., Kusakunniran, W., Wu, Q., Xu, J. & Zhang, J. Recognizing Gaits Across Walking and Running Speeds. *ACM Transactions on Multimedia Computing, Communications, and Applications* **18,** 1–22. https://doi.org/10.1145/3488715 (Aug. 2022).

104. Yu, S., Tan, D. & Tan, T. *A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition* in *Proceedings - International Conference on Pattern Recognition* **4** (IEEE, 2006), 441–444. https://doi.org/10.1109/ICPR.2006.67.

105. Zhang, Y., Huang, Y., Yu, S. & Wang, L. Cross-view gait recognition by discriminative feature learning. *IEEE Transactions on Image Processing* **29,** 1001–1015. https://doi.org/10.1109/TIP.2019.2926208 (2020).

106. Zhang, Z., Tran, L., Liu, F. & Liu, X. On Learning Disentangled Representations for Gait Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **44,** 345–360. https://doi.org/10.1109/TPAMI.2020.2998790 (Jan. 2022).

107. Zheng, J. *et al. Gait Recognition in the Wild with Multi-hop Temporal Switch* in *Proceedings of the 30th ACM International Conference on Multimedia* (ACM, Oct. 2022), 6136–6145. https://doi.org/10.1145/3503161.3547897.

108. Zou, Y., He, N., Sun, J., Huang, X. & Wang, W. Occluded Gait Emotion Recognition Based on Multi-Scale Suppression Graph Convolutional Network. *Computers, Materials Continua* **82,** 1255–1276. https://doi.org/10.32604/CMC.2024.055732 (Jan. 2025).