QAGait: Revisit Gait Recognition from a Quality Perspective

Zengbin Wang[1*]**, Saihui Hou**[2,3*]**, Man Zhang**[1†]**, Xu Liu**[3]**,**
Chunshui Cao[3]**, Yongzhen Huang**[2,3]**, Peipei Li**[1]**, Shibiao Xu**[1]

[1]Beijing University of Posts and Telecommunications
[2]Beijing Normal University
[3]Watrix AI
{wzb1, zhangman, lipeipei, shibiaoxu}@bupt.edu.cn,
{housaihui, huangyongzhen}@bnu.edu.cn, {xu.liu, chunshui.cao}@watrix.ai

Abstract

Gait recognition is a promising biometric method that aims to identify pedestrians from their unique walking patterns. Silhouette modality, renowned for its easy acquisition, simple structure, sparse representation, and convenient modeling, has been widely employed in controlled in-the-lab research. However, as gait recognition rapidly advances from in-the-lab to in-the-wild scenarios, various conditions raise significant challenges for silhouette modality, including 1) unidentifiable low-quality silhouettes (abnormal segmentation, severe occlusion, or even non-human shape), and 2) identifiable but challenging silhouettes (background noise, non-standard posture, slight occlusion). To address these challenges, we revisit gait recognition pipeline and approach gait recognition from a quality perspective, namely QAGait. Specifically, we propose a series of cost-effective quality assessment strategies, including Maxmial Connect Area and Template Match to eliminate background noises and unidentifiable silhouettes, Alignment strategy to handle non-standard postures. We also propose two quality-aware loss functions to integrate silhouette quality into optimization within the embedding space. Extensive experiments demonstrate our QAGait can guarantee both gait reliability and performance enhancement. Furthermore, our quality assessment strategies can seamlessly integrate with existing gait datasets, showcasing our superiority. Code is available at https://github.com/wzb-bupt/QAGait.

1 Introduction

Gait recognition has attracted widespread interest in computer vision and biometrics community due to its potential to identify individuals from a distance. Unlike other biometric characteristics such as face, fingerprint, and iris, gait is difficult to disguise and can be non-intrusively acquired without any cooperation (Sepas-Moghaddam and Etemad 2022). These inherent characteristics contribute to gait recognition a promising solution for wide applications, including surveillance, suspect tracking, motion monitoring, and identification (Wan, Wang, and Phoha 2018; Shen et al. 2022).

One notable advantage of gait recognition, as highlighted in previous research (Fan et al. 2023; Ma et al. 2023; Wang

---

*These authors contributed equally to this work.

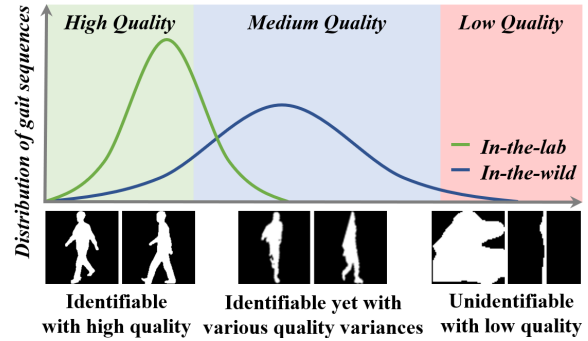†Corresponding author (zhangman@bupt.edu.cn).

Figure 1: The quality difference when advanced from in-the-lab to in-the-wild gait datasets, including: 1) more extremely low-quality silhouettes that are unidentifiable; and 2) higher quality variances due to various and complex environments.

et al. 2023a), is its common use of binary silhouette as input. This preference stems from its simple structure and sparse representation, which allows researchers to focus on extracting essential motion patterns and action variations, rather than being distracted by background, color, or texture details from RGB images. Moreover, this simplified input enables efficient and practical processing. Extensive in-the-lab experiments with shallow models have achieved impressive performance, *i.e.*, 95.9% on CASIA-B dataset (Hsu et al. 2022) and 92.1% on OU-MVLP dataset (Dou et al. 2023).

In most cases, the immense success of in-the-lab gait datasets can be attributed to their high-quality segmentation, controlled viewpoint of standard posture, minimal occlusion, *etc*. Under these controlled settings, a simple data preprocessing approach (Chao et al. 2019; Fan et al. 2023) can easily ensure that pedestrians are uniformly processed, thus facilitating subsequent feature extraction and analysis.

However, as in Figure 1, when gait recognition advances from controlled in-the-lab to in-the-wild scenarios (*e.g.*, Gait3D (Zheng et al. 2022) and GREW (Zhu et al. 2021)), it encounters various challenges such as complex background, arbitrary camera position, and crowded environment. These uncontrolled conditions lead to two key issues:

- *Unidentifiable with low quality*: The traditional data preprocessing approach mentioned above is insufficient to

handle the complexities of uncontrolled conditions, potentially reserving a significant number of unexpected challenging silhouettes. Even though some are unidentifiable, all of them are passed to the gait model for feature learning and may potentially affect the model prediction.

- *Identifiable yet with various quality variances*: Considerable noise and variability observed across multiple environments increase the uncertainty of gait sequence, resulting in various quality variances. However, current methods usually treat all gait sequences uniformly.

These two issues potentially mislead gait model to learn biased gait patterns, thereby resulting in poor performance when compared with in-the-lab settings. In this case, a key question arises: Is there a *cost-effective quality assessment* or *quality-aware* approach to mitigate the negative impact of background noises, unidentifiable silhouettes, and identifiable silhouettes yet with various quality variances?

In this paper, we revisit gait recognition from a quality perspective and propose a *unified quality assessment* and *quality-aware* framework, *i.e.*, **QAGait**, to address the above challenges in a *cost-effective* way. Specifically, before feeding into the feature extractor, we propose a series of quality assessment strategies to guarantee silhouette quality, including Maximal Connect Area, Template Match, and Alignment. In the feature learning stage, we introduce a gait quality indicator and propose two quality-aware loss functions to adaptively adjust optimization based on inherent gait quality, including Quality Adaptive Margin CE loss (QACE) and Quality Adaptive Margin Triplet loss (QATriplet).

- To the best of our knowledge, we make one of the first attempts to explore *quality-oriented* gait recognition, including unified quality assessment strategies and quality-aware feature learning.

- Our thoughtful quality assessment strategies provide a unified and cost-effective approach, suitable for both in-the-wild and in-the-lab gait datasets.

- Our novel quality-aware feature learning incorporates a reliable gait quality indicator and two quality-aware loss functions, allowing us to dynamically adjust the optimization progress from a quality perspective.

- Extensive experimental results on in-the-wild and in-the-lab gait datasets demonstrate the effectiveness of our method. For example, we achieve a remarkable 7.3% improvement in Rank-1 accuracy on Gait3D dataset.

2 Related Work

2.1 Gait Recognition

Currently, gait recognition can be primarily categorized into model-based and appearance-based methods.

Model-based methods mostly utilize 2D/3D pose, 3D mesh, and point cloud as input. For example, GaitGraph and GaitGraph2 (Teepe et al. 2021, 2022) utilize human pose estimation to extract 2D poses from RGB images and combine GCN for spatial-temporal modeling. SMPLGait (Zheng et al. 2022) employs a 3D SMPL model to learn 3D parameters of body shape, pose, and camera viewpoint as supplementary information. LidarGait and LiCamGait (Shen et al.

2023; Han et al. 2022) leverage LiDAR to explore point cloud for gait recognition. However, all these methods, along with others not explicitly mentioned (Pinyoanuntapong et al. 2023; Wu et al. 2023) rely on the prediction from estimation models of pose, mesh, point cloud, and others.

Appearance-based methods are widely used to learn the inherent spatial and temporal variations in body shape, clothing, and movement dynamics. GEINet (Shiraga et al. 2016) aggregates each gait sequence to a gait energy image (GEI) for feature learning. GaitSet (Chao et al. 2019, 2021) regards each gait sequence as an unordered set and lots of works (Hou et al. 2020, 2021; Chai et al. 2021) follow this setting. GaitPart (Fan et al. 2020) proposes to improve part-based feature learning from focal convolution. GaitGL (Lin, Zhang, and Yu 2021) develops a 3D global-local feature extractor to ensemble global and local features. CSTL (Huang et al. 2021a) achieves adaptive temporal learning and salient spatial mining. MetaGait (Dou et al. 2022a) proposes to capture omni-scale dependency from spatial/channel/temporal dimensions of gait sequences. DyGait (Wang et al. 2023a) proposes to focus on the extraction of dynamic features. Furthermore, other methods such as 3D-Local (Huang et al. 2021b), LagrangeGait (Chai et al. 2022), GaitMPL (Dou et al. 2022b), GaitGCI (Dou et al. 2023), and DANet (Ma et al. 2023) are continuously emerging due to the manifold advantages of using silhouettes, such as easy acquisition, simple structure, sparse representation, and convenient modeling. Meanwhile, as gait recognition moves to real-world scenarios, gait quality becomes crucial, and some researches now require high-quality silhouette (Wang et al. 2023b).

2.2 Image Quality Assessment

In computer vision, Image Quality Assessment (IQA) is crucial for evaluating the inherent quality of images. Various methods can be broadly categorized into two types: model-based and model-free approaches. Model-based methods (Yang et al. 2022; Madhusudana et al. 2022; Pan et al. 2022) utilize neural networks to automatically learn image quality features. Model-free approaches (Al-Najjar and Chen 2012; Nascimento et al. 2023; Abdelreheem et al. 2023) usually employ well-designed perceptual metrics such as the Structural Similarity Index (SSIM) and Peak Signal-to-Noise Ratio (PSNR) to quantify image quality. Additionally, various techniques available in libraries like OpenCV can be utilized for image quality enhancement, such as image smoothing, denoising, histogram equalization, *etc*.

However, quality assessment in gait task is underdeveloped. GQAN (Hou et al. 2022) attempts to evaluate gait quality using network design at frame and part levels to emphasize interpretability. Differently, our method seeks to mitigate challenges from unidentifiable and variable-quality silhouettes through *quality assessments* and *quality-aware feature learning* without network modification.

3 Methodology

3.1 Revisit Gait Recognition Pipeline

1) Early Data Processing. Existing appearance-based gait models often apply a simple data preprocessing ap-
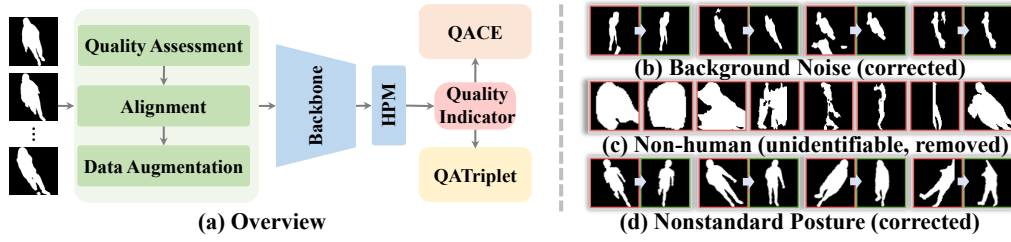
Figure 2: Overview of our QAGait and some examples with quality issues. (a) We design a series of cost-effective quality assessment strategies, alignment, and data augmentation for given gait sequences before feeding them into backbone. During feature learning, we introduce a quality indicator and propose two quality-aware loss functions based on CE loss and Triplet loss (*i.e.*, QACE and QATriplet) to improve model adaptability for various quality variances between gait sequences. (b-d) Some examples with diverse quality issues from in-the-wild gait datasets. HPM: Horizontal Pyramid Mapping (Chao et al. 2019).

proach (Chao et al. 2019; Fan et al. 2023) to process both in-the-lab and in-the-wild datasets, including:

- Crop out black regions at top of head and bottom of feet.
- Resize image to the expected height.
- Horizontally move human to image center and crop extra side regions to the expected width.

[TODO] However, this process is conducted from early gait recognition and tailored for controlled in-the-lab gait datasets. When transmitting to complex in-the-wild datasets, it is not always suitable for in-the-wild datasets with complex backgrounds, non-human shapes, nonstandard postures, and other unforeseen scenarios.

2) Data Augmentation. Data augmentation in gait recognition is an effective approach to expand the diversity of training dataset, enhance the generalization capability of the model, and prevent overfitting, including:

- Horizontal Flip (HF), Rotation (R), Perspective Transformation (PT), Affine Transformation (AT), Random Erasing (RE).

[TODO] However, we note that nonstandard postures are quite common in outdoor datasets (Zheng et al. 2022) due to their unnormalized camera perspectives. This makes using Rotation or Perspective Transformation on an extremely tilted person (*e.g.*, the line connecting the head and feet is nearly parallel to the ground) potentially create even worse patterns that are unidentifiable.

3) Feature Extraction. During the feature extraction stage, the goal is to transform the preprocessed gait sequences into discriminative and compact feature representations.

Given that this paper does not primarily emphasize the backbone, as a trade-off, we select the latest GaitBase (Fan et al. 2023) as our base model for feature representation.

4) Loss Function. In the embedding space, the *softmax cross-entropy loss* and *triplet loss* serve as fundamental loss functions in current gait models (Fan et al. 2023). The former measures the discrepancy between predicted class probabilities and true labels, while the latter pulls the positive pair closer and pushes the negative pair further.

[TODO] However, both of them treat all gait sequences equally and do not account for the significant uncertainty associated with various quality variances in different gait sequences. This oversight severely impacts both the convergence speed and its discrimination capability.

3.2 Model Overview

As shown in Figure 2, our QAGait presents a comprehensive approach for quality modeling. In quality assessments, we start from Maximal Connect to remove background noise. Only then can we introduce Template Match for accurate shape matching, leaving normal silhouettes for improved Lean-aware Alignment. On top of that, we observe that high quality variance still exists. Thus we further introduce quality-aware losses to mitigate the impact of remaining low-quality silhouettes, which is agnostic to the network and computationally efficient.

3.3 Unified Silhouette Quality Assessment

In this section, as shown in Figure 3(a), we introduce two unified quality assessment strategies, *i.e.*, Maxmial Connect Area and Template Match, aiming to remove the quality issues mentioned above in a *cost-effective* manner. We expect the remaining gait silhouettes can preserve the unique gait patterns, not affected by unrelated factors.

1) Maximal Connect Area. Since gait recognition is frequently encountered in complex backgrounds and crowded environments, some obvious segmentation errors often exist in outdoor gait datasets. This kind of segmentation error can be largely regarded as background noise and always dissociates from the human body. Thus, as illustrated in Figure 3(a), we develop a Maxmial Connect Area process as our first quality assessment strategy to remove these obvious mistakes.

For details, we calculate the Maximal Connect Area ($area$) by OpenCV (*i.e.*, $cv2.connectedComponents(\cdot)$) for each silhouette, and set a threshold (ϵ) for two cases:

- When $area \geq \epsilon$ and small regions exist, we simply remove these small regions and reserve the main body.
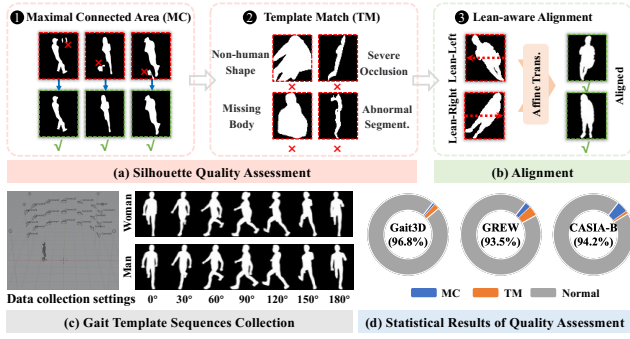- When $area < \epsilon$, the whole silhouette will be removed.

Figure 3: Top: Examples of our silhouette quality assessment and alignment. Bottom: Our gait template sequence collection and the statistical results of quality assessment.

2) Template Match. As shown in Figure 3(a), only when the first Maximal Connect Area progress removes background noise can we proceed with the second Template Match progress to further remove unidentifiable silhouettes, such as the person with non-human shape, severe occlusion, missing body, or abnormal segmentation.

Building on remarkable processes in 3D SMPL (Loper et al. 2023) and improving industry standards in gait recognition (Mirelman et al. 2018), leveraging their well-established frameworks makes it possible to create a credible benchmark for standard gait template sequences based on 3D virtual human. To this end, as shown in Figure 3(c), we construct an optimal environment and collect a series of standard gait sequences, including various conditions:

- Multiple horizontal views: $0°$-$180°$, with a $30°$ interval.
- Multiple vertical heights: 1.5 m, 2.5 m, and 3.5 m.
- Different genders: woman and man.

Even though these standard gait template sequences are available, two challenges still persist for our template match: 1) Non-normalized camera views in outdoor scenes cause deviations (rotation) from the standard pose; and 2) Varying camera distances lead to scale differences. Based on these two observations, we employ *Hu Moments*[1], renowned for their robustness against rotation and scale variations, as our shape-matching algorithm and set a low threshold (τ) to effectively remove non-human shapes.

3.4 Lean-aware Data Augmentation and Alignment

As mentioned above, we consistently emphasize the non-standard postures observed in unnormalized camera views. In this case, the commonly used Rotation (R) and Perspective Transformation (PT) in data augmentation may result in challenging and even unidentifiable samples. Our analysis indicates that current methods lack sensitivity to lean human body. For example:

---

[1]Hu moments contain seven distinct invariant moments that exhibit invariance to translation, rotation, and scale changes, providing a compact representation to describe images. This can be efficiently achieved by $cv2.matchShapes(\cdot)$ in OpenCV.

- A person who leans significantly to the left (*e.g.*, $60°$ or more) is not suitable to apply leftward rotation.

To address this, we introduce a strict constraint (*i.e.*, lean-aware) in R and PT to prevent excessive deviations. Furthermore, to comprehensively tackle this issue, we propose a lean-aware alignment method aimed at rectifying all silhouettes with nonstandard postures.

1) Lean-aware Rotation and Perspective Transformation. To begin with, we uniformly divide the human body into four regions: top-left, top-right, bottom-left, and bottom-right. We then compare the calculated area of each region:

- If top-left area is larger than top-right by a certain proportion (*e.g.*, 50%), and bottom-left is smaller than bottom-right by a certain proportion, it is denoted as "*Lean-Left*", avoiding leftward rotation.
- The same to "*Lean-Right*", without rightward rotation.

2) Lean-aware Alignment. To fundamentally address the lean body in gait sequences and enable consistent learning, as shown in Figure 3(b), we introduce a lean-aware alignment module, including four steps as follows.

- **Step1**: Find minimal bounding box[2] around human body.
- **Step2**: Calculate rotation angle between the longer side and vertical line.
- **Step3**: Generate affine transformation for correction.
- **Step4**: Add a slight rotation disturbance (θ) to aligned gait sequence for diversity.

It is worth noting that a positive or negative angle indicates the lean direction, so we call it Lean-aware Alignment. In our implementation, we carefully compare *frame-level* and *sequence-level* alignment and select the latter (*i.e.*, average angles across all silhouettes in a gait sequence.) to ensure consistency within a gait sequence.

3.5 Quality-aware Feature Learning

In this section, we establish a connection between gait reliability and quality to mitigate various quality variances in remaining gait sequences. Specifically, we introduce a *Quality Indicator* for gait quality assessment and update standard Cross-Entropy loss and Triplet loss toward *Quality Adaptive Margin CE loss (QACE)* and *Quality Adaptive Margin Triplet loss (QATriplet)* for quality-aware feature learning.

1) Quality Indicator. Given the pivotal role of partial feature representation in gait recognition (*e.g.*, horizontal strips division to pursue more clues for human body (Chao et al. 2019; Fan et al. 2020)) and the compatibility of feature norm within model-free image quality assessment approaches (Li et al. 2018; Schlett et al. 2022; Kim, Jain, and Liu 2022), we adopt *Partial Feature Norm* as our quality indicator.

$$\widehat{||p_i||} = \left\lfloor \frac{\overline{||p_i||} - \mu_p}{\sigma_p/h} \right\rceil_{-1}^{1} \tag{1}$$

---

[2]The minimal bounding box is strongly related to the above quality assessment strategies for that background noise and non-human shape will affect the bounding selection and angle calculation. This can be efficiently achieved by $cv2.minAreaRect(\cdot)$.

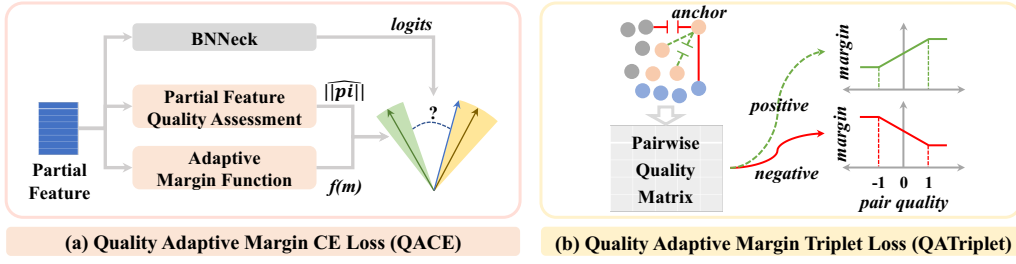**(a) Quality Adaptive Margin CE Loss (QACE)**   **(b) Quality Adaptive Margin Triplet Loss (QATriplet)**

Figure 4: Our quality-aware loss functions facilitate the optimization of trustworthy gait sequences and mitigate the effect of low-quality gait sequences. Left: QACE utilizes Partial Feature Quality as the quality indicator to generate adaptive margin functions. This is then applied to the logits of the model output. Right: QATriplet emphasizes Pairwise Quality as the quality indicator for adaptively adjusting positive and negative pair margins, respectively. BNNeck (Luo et al. 2019).

where p_i indicates partial features extracted from the backbone, $\overline{||p_i||}$ indicates average partial features, μ_p and σ_p denote the batch mean and variance, $\lfloor \cdot \rceil_{-1}^{1}$ constrains values to facilitate subsequent processing (*i.e.*, $\widehat{||p_i||} \in [-1, 1]$, from low quality to high quality), h indicates that we follow "3-σ rule" (also known as the "68-95-99.7 rule") in normal distribution and set $h = 0.33$ (Lehmann 2013).

Discussion. The quality indicator is a soft manner to integrate quality into loss calculation to mitigate the effect of low-quality frames. Its necessity comes from two aspects: 1) Real-world gait data varies in quality, making it hard to adopt a fixed threshold to distinguish high- or low-quality frames easily. Moreover, highly reliable quality metrics are still lacking. 2) To some extent, the low-quality frames (but identifiable) are necessary to ensure the generalization of gait recognition to multiple scenes (*e.g.*, low-resolution).

2) Quality Adaptive Margin CE loss (QACE).   The softmax cross-entropy loss is widely used for classification and has been advanced to margin-based forms like ArcFace (Deng et al. 2019) and CosFace (Wang et al. 2018). The introduction of a margin can compress the learning space to enhance the discriminative capability of the model. Here is the combined form of both as follows.

$$L_1 = -\frac{1}{N} \sum_{i=1}^{N} \log \frac{e^{s(\cos(\theta_{y_i} + m_{angle}) - m_{add})}}{e^{s(\cos(\theta_{y_i} + m_{angle}) - m_{add})} + \sum_{j=1, j \neq y_i}^{N} e^{s \cos \theta_{y_i}}} \quad (2)$$

where m_{angle} and m_{add} are angular and additive margin for ArcFace and CosFace, s indicates scale transformation, y_i is the true labels. The learned embedding features are distributed on a hypersphere with a radius of s.

However, we notice that a fixed margin always treats all features uniformly even when they have varying quality distributions. This will lead to: 1) Lower-quality gait sequences often cluster near the classification boundary since its mixed and unstable information, and the margin may have a negative effect. 2) Higher-quality gait sequences tend close to the classification center, allowing for a larger margin to compress feature space further.

Therefore, we combine the quality indicator into the margin and dynamically adjust the margin based on the quality of gait sequence. Empirically, we propose to calculate the

corresponding m_{angle} and m_{add} as follows.

$$m_{angle} = -m_1 \cdot \widehat{||p_i||}, \quad m_{add} = m_1 \cdot \widehat{||p_i||} + m_1 \quad (3)$$

where m_1 is the initial margin and $\widehat{||p_i||} \in [-1, 1]$ serves as the quality indicator. When $\widehat{||p_i||} \to \{-1\}$, the overall margin is a simple m_{angle} with initial m_1. When $\widehat{||p_i||} \to \{1\}$, m_{angle} and m_{add} all tend to assign a larger value.

3) Quality Adaptive Margin Triplet loss (QATriplet). Triplet loss primarily focuses on pairwise samples, pulling the positive pair closer and pushing the negative pair further apart. The original triplet loss can be simply formulated as:

$$L_2 = \max \{d(a, p) - d(a, n) + m_2, 0\} \quad (4)$$

Similar to QACE, the original triplet loss treats all positive or negative pairs equally. However, we argue that optimizing gait model with lower-quality pairs may not always yield reliable results. For higher-quality pairs, a more effective strategy is to increase the optimization distance for positive pairs while reducing the optimization distance for negative pairs.

$$L_3 = \max \{(d(a, p) + m_{ap}) - (d(a, n) - m_{an}), 0\} \quad (5)$$

We define the pairwise quality indicator ($PQ(s_1, s_2)$) and propose to adaptively adjust the margin (m_{ap}, m_{an}) based on pairwise quality for an appropriate optimization process.

$$PQ(s_1, s_2) = \min \left\{ \widehat{||p_{i, s_1}||}, \widehat{||p_{i, s_2}||} \right\} \quad (6)$$

where $\min$ indicates that if one of gait sequences is unreliable, the overall pairwise quality becomes compromised.

$$m_{ap} = m_{an} = 1.5 \cdot m_2 + (0.5 \cdot m_2 \cdot PQ(s_1, s_2)) \quad (7)$$

where m_2 is the initial margin, s_1 and s_2 are pairwise gait sequences. m_{ap} and m_{an} are personalized margins for positive and negative pairs. The coefficients (1.5 and 0.5) are used to confine the range of m_{ap} or m_{an} within $[m_2, 2m_2]$. $PQ \in [-1, 1]$. As $PQ \to \{-1\}$, a minimal distance of m_2 is added, while as $PQ \to \{1\}$, a maximal distance of $2m_2$ is added. These correspond to small and large optimization distances, respectively.

0. Original Benchmark (Gait3D | GREW)

Method	Source	Gait3D | Gait3D$^{\#}$			GREW | GREW$^{\#}$			
		R-1	R-5	mAP	R-1	R-5	R-10	R-20
GaitSet	AAAI'19	36.7	58.3	30.0	46.3	63.6	70.3	76.8
GaitPart	CVPR'20	28.2	47.6	21.6	44.0	60.7	67.3	73.5
GaitGL	ICCV'21	29.7	48.5	22.3	47.3	63.6	69.3	74.2
CSTL	ICCV'21	11.7	19.2	5.6	50.6	65.9	71.9	-
SMPLGait	CVPR'22	46.3	64.5	37.2	-	-	-	-
MTSGait	ACM MM'22	48.7	67.1	37.6	55.3	71.3	76.9	81.6
DAGait	CVPR'23	48.0	69.7	-	-	-	-	-
GaitBase	CVPR'23	62.4	77.7	52.3	58.9	73.7	79.1	83.1

1. After Quality Assessment (Gait3D$^{\#}$ | GREW$^{\#}$)

Method	Source	Gait3D | Gait3D$^{\#}$			GREW | GREW$^{\#}$			
GaitSet	AAAI'19	$41.7^{\uparrow5.0}$	$61.5^{\uparrow3.2}$	$32.6^{\uparrow2.6}$	$46.5^{\uparrow0.2}$	$64.0^{\uparrow0.4}$	$70.5^{\uparrow0.2}$	$76.9^{\uparrow0.1}$
GaitPart	CVPR'20	$31.3^{\uparrow3.1}$	$50.7^{\uparrow3.1}$	$23.9^{\uparrow2.3}$	$46.3^{\uparrow2.3}$	$63.2^{\uparrow2.5}$	$69.5^{\uparrow2.2}$	$74.4^{\uparrow0.9}$
GaitGL	ICCV'21	$34.1^{\uparrow4.4}$	$53.4^{\uparrow4.9}$	$25.1^{\uparrow2.8}$	$51.4^{\uparrow4.1}$	$66.8^{\uparrow3.2}$	$72.4^{\uparrow3.1}$	$76.6^{\uparrow2.4}$
GaitBase	CVPR'23	$59.7^{\downarrow2.7}$	$77.6^{\downarrow0.1}$	$51.0^{\downarrow1.3}$	$57.2^{\downarrow1.7}$	$72.3^{\downarrow1.4}$	$77.1^{\downarrow2.0}$	$81.3^{\downarrow1.8}$

2. After Quality Assessment & Quality-aware (Gait3D$^{\#}$ | GREW$^{\#}$)

Method		Gait3D | Gait3D$^{\#}$			GREW | GREW$^{\#}$			
QAGait		$67.0^{\uparrow7.3}$	$81.5^{\uparrow3.9}$	$56.5^{\uparrow5.5}$	$59.1^{\uparrow1.9}$	$74.0^{\uparrow1.7}$	$79.2^{\uparrow2.1}$	$83.1^{\uparrow1.8}$

Table 1: Comparison results on Gait3D and GREW.

0. Original Benchmark (CASIA-B)

Method	Source	NM	BG	CL
GaitSet	AAAI'19	95.0	87.2	70.4
GaitPart	CVPR'20	96.2	91.5	78.7
MT3D	MM'20	96.6	92.9	82.2
GaitGL	ICCV'21	97.4	94.5	83.6
3DLocal	ICCV'21	97.5	94.3	83.7
CSTL	ICCV'21	97.8	93.6	84.2
LagrangeGait	CVPR'22	97.5	94.6	85.1
MetaGait	ECCV'22	98.1	95.2	86.9
DANet	CVPR'23	98.0	95.9	89.9
†GaitBase	CVPR'23	96.7	93.0	75.1

1. After Quality Assessment (CASIA-B$^{\#}$)

Method	Source	NM	BG	CL
†GaitBase	CVPR'23	$97.8^{\uparrow1.1}$	$94.3^{\uparrow1.3}$	$77.1^{\uparrow2.0}$

2. After Quality Assessment & Quality-aware (CASIA-B$^{\#}$)

Method		NM	BG	CL
†QAGait		$97.9^{\uparrow0.1}$	$94.6^{\uparrow0.3}$	$78.2^{\uparrow1.1}$

Table 2: Comparison results on CASIA-B, $\dagger$ indicates we reduce the last two layers to suit small-scale CASIA-B.

4 Experiments

4.1 Experimental Settings

Datasets. Our method is primarily evaluated on two in-the-wild gait datasets, Gait3D (Zheng et al. 2022) and GREW (Zhu et al. 2021), due to their complex data collection environments and various covariates in outdoor scenes. We also include CASIA-B (Yu, Tan, and Tan 2006) since it utilizes outdated background subtraction for segmentation. We strictly follow the original training and test settings.

Training Details. We adopt the latest GaitBase (Fan et al. 2023) as our backbone. All data are normalized to 64×44. We randomly select P identities and corresponding K sequences in a mini-batch (*i.e.*, {32,4} for Gait3D / GREW and {8,16} for CASIA-B). Each sequence contains 30 randomly sampled silhouettes. We use SGD optimizer with weight decay $5e$-4. The initial learning rate is 0.1 and decays

10 times at each milestone (*i.e.*, {20K,40K,50K} for Gait3D / CASIA-B, and {80K,120K,150K} for GREW). The total iterations are 60K for Gait3D / CASIA-B and 180K for GREW. The overall loss function is: $L = L_1 + L_3$.

Implementation Details. For a fair comparison, we retain at least 15 frames for a sequence when too many frames need to be removed in the quality assessment step. For the margin setting, we apply grid search and select the optical ($m_1 = 0.1$, $s = 8$) for ArcFace (Table 6), and our QACE follows this setting. For QATriplet, we follow the latest gait research and set $m_2 = 0.15$ to achieve the average margin of m_{ap} and m_{an} is around 0.2. The threshold[3] in Maximal Connect Area and Template Match is $\epsilon = 0.95$ and $\tau = 0.001$. The random disturbance in Alignment is set as $\theta = 5°$.

4.2 Comparison with State-of-The-Art Methods

In this section, we evaluate our method on two aspects: 1) original benchmark with our quality assessment (denoted as Gait3D$^{\#}$/GREW$^{\#}$/CASIA-B$^{\#}$), and 2) original benchmark with our whole QAGait (*i.e.*, quality assessment, alignment, data augmentation, and quality-aware feature learning).

Evaluation on outdoor datasets (Gait3D & GREW). As shown in Figure 3(d) and Table 1, our quality assessment strategies on original Gait3D and GREW can effectively remove unidentifiable silhouettes and reduce noise impact to some extent, *i.e.*, removing 3.2% / 6.5% silhouettes for Gait3D / GREW. This yields remarkable improvements based on current shallow models. Take GaitPart / GaitGL for example, we achieve 3.1% / 4.4% and 2.3% / 4.1% Rank-1 gains for Gait3D$^{\#}$ and GREW$^{\#}$ from their original settings. These enhancements align well with our motivation.

We also adopt the latest GaitBase while it experiences minor drops under default settings (*e.g.*, 62.4%→59.7% in

---

[3]We invite 20 volunteers, half from our group and half from the college. Each is assigned 30 sequences and asked to select thresholds to distinguish identifiable and unidentifiable frames. Finally, we set the threshold by calculating the *median* of their selections.

Conditions	R-1	R-5	mAP
Original Gait3D (w/o DA)	55.1	72.8	46.6
+QA(MaxConnect)	55.8	72.9	45.7
+QA(MaxConnect+TemplateMatch)	54.9	73.0	45.9
+QA+DA	62.2	78.7	51.6
+QA+DA+Align	63.5	79.8	54.2
+QA+DA+Align+QACE	66.0	80.5	56.2
+QA+DA+Align+QACE+QATriplet	**67.0**	**81.5**	**56.5**

Table 3: Evaluation results on each component in QAGait.

HF	R	PT	AT	RE	R-1	R-5	mAP
✓	✓	✓			59.7	77.6	51.0
✓	✓(Lean)	✓(Lean)			61.1	77.2	50.8
✓	✓(Lean)			✓	**62.2**	**78.7**	**51.6**

Table 4: Ablation study on Data Augmentation. "Lean" refers to restricting rotation or perspective transformation.

Gait3D$^{\#}$). The probable reason is the mismatch of original data augmentation with the corrected data. The performance can be improved with our data augmentation (59.7%→62.2% in Table 4). Additionally, our cost-effective quality-aware feature learning mitigates various quality issues and achieves remarkable enhancements of 7.3% and 1.9% Rank-1 accuracy on Gait3D$^{\#}$ and GREW$^{\#}$ datasets, highlighting the substantial utility of our approach.

Evaluation on CASIA-B. Since CASIA-B dataset uses outdated background subtraction for segmentation, some segmentation errors may exist. Nevertheless, unlike the Gait3D and GREW datasets, CASIA-B is meticulously collected within a controlled in-the-lab environment, thereby preventing error amplification. Table 2 shows that our quality assessment yields a modest 2.0% improvement on CL condition based on GaitBase, and our whole QAGait can further provide an additional 1.1% improvement.

4.3 Ablation Study

Effectiveness of different components in QAGait. As shown in Table 3, we present our overall approach encompassing quality assessment (QA), alignment, and quality-aware feature learning (QACE & QATriplet). Our QAGait can progressively improve recognition performance.

Impact of strict constraint in data augmentation. As shown in Table 4, lean-aware R and PT can enhance data credibility and lead to a 1.4% Rank-1 improvement. Moreover, our exploration also reveals a new data augmentation combination (*i.e.*, HF, R, and RE) that is better suitable for Gait3D$^{\#}$, with an additional 1.1% recognition enhancement.

Frame-level	Sequence-level	Disturb	R-1	R-5	mAP
-	-	-	62.2	78.7	51.6
✓			63.0	78.9	53.2
✓		✓	63.0	79.1	53.7
	✓	✓	**63.5**	**79.8**	**54.2**

Table 5: Ablation study on alignment types and disturbance.

Fixed Margin	Quality-aware Margin	Quality Calculation	R-1	R-5	mAP
-	-	-	63.5	79.8	54.2
✓(0.1)			65.0	80.2	55.6
	✓	Sum-Partial	65.3	80.8	55.4
	✓	Min-Partial	64.0	**80.5**	55.1
	✓	Avg-Partial	**66.0**	**80.5**	**56.2**

Table 6: Ablation study on quality-aware margin and quality calculation forms in QACE.

Fixed Margin	Pair Quality Margin	Pair Quality Calculation	R-1	R-5	mAP
✓(0.2)			66.0	80.5	56.2
	✓	AvgQualityPair	63.9	80.7	54.2
	✓	MinQualityPair	**67.0**	**81.5**	**56.5**

Table 7: Ablation study on pairwise quality adaptive margin and pairwise quality calculation forms in QATriplet.

Impact of align types in alignment. We experiment with Frame / Sequence-level Alignment and explore the impact of random disturbance. Table 5 shows that sequence-level alignment with a 5° random disturbance achieves higher performance, indicating that alignment facilitates unified perspective learning and perturbations enhance model diversity.

Impact of quality adaptive margin in QACE. Here we introduce margin into CE loss and develop a quality-aware adaptive margin CE loss (QACE) for evaluation. Table 6 reveals that a fixed margin (*i.e.*, $m = 0.1$, similar to ArcFace) enhances Rank-1 accuracy by 1.5%. Moreover, our QACE, with average partial quality, provides an additional performance gain, surpassing the original CE loss by 2.5%.

Impact of pairwise quality calculation in QATriplet. Our QATriplet emphasizes pairwise quality of gait sequences and dynamically adjusts optimization distance. Table 7 shows that our QATriplet achieves 1.0% performance gain with minimizing pairwise quality, highlighting the adaptability to diverse gait sequences.

Discussion about computation cost. Our approach is computationally efficient largely due to two aspects. 1) The quality assessments are customized for gait based on low-level image processing (OpenCV), making it both affordable and readily available. 2) The quality-aware feature learning is agnostic to the network and non-parametric with only minimal additional computation (<1% training time). Furthermore, the inference time is unaffected.

5 Conclusion

In conclusion, we present a comprehensive study of gait recognition pipeline and propose a novel quality-oriented gait recognition, namely QAGait. Our method involves cost-effective quality assessment strategies and quality-aware feature learning to effectively address challenges arising from unidentifiable silhouettes and identifiable silhouettes yet with various quality variances. We expect our method to provide a convenient way to assess gait recognition quality and inspire future research in quality-aware gait recognition.

Acknowledgments

References

Abdelreheem, A.; Eldesokey, A.; Ovsjanikov, M.; and Wonka, P. 2023. Zero-Shot 3D Shape Correspondence. *arXiv preprint arXiv:2306.03253*.

Al-Najjar, Y.; and Chen, D. 2012. Comparison of image quality assessment: PSNR, HVS, SSIM, UIQI. *International Journal of Scientific and Engineering Research*, 3(8): 1–5.

Chai, T.; Li, A.; Zhang, S.; Li, Z.; and Wang, Y. 2022. Lagrange motion analysis and view embeddings for improved gait recognition. In *CVPR*, 20249–20258.

Chai, T.; Mei, X.; Li, A.; and Wang, Y. 2021. Silhouette-based view-embeddings for gait recognition under multiple views. In *ICIP*, 2319–2323. IEEE.

Chao, H.; He, Y.; Zhang, J.; and Feng, J. 2019. GaitSet: Regarding gait as a set for cross-view gait recognition. In *AAAI*, volume 33, 8126–8133.

Chao, H.; Wang, K.; He, Y.; Zhang, J.; and Feng, J. 2021. GaitSet: Cross-view gait recognition through utilizing gait as a deep set. *TPAMI*, 44(7): 3467–3478.

Deng, J.; Guo, J.; Xue, N.; and Zafeiriou, S. 2019. Arcface: Additive angular margin loss for deep face recognition. In *CVPR*, 4690–4699.

Dou, H.; Zhang, P.; Su, W.; Yu, Y.; and Li, X. 2022a. Metagait: Learning to learn an omni sample adaptive representation for gait recognition. In *ECCV*, 357–374. Springer.

Dou, H.; Zhang, P.; Su, W.; Yu, Y.; Lin, Y.; and Li, X. 2023. GaitGCI: Generative counterfactual intervention for gait recognition. In *CVPR*, 5578–5588.

Dou, H.; Zhang, P.; Zhao, Y.; Dong, L.; Qin, Z.; and Li, X. 2022b. GaitMPL: Gait recognition with memory-augmented progressive learning. *TIP*.

Fan, C.; Liang, J.; Shen, C.; Hou, S.; Huang, Y.; and Yu, S. 2023. OpenGait: Revisiting gait recognition towards better practicality. In *CVPR*, 9707–9716.

Fan, C.; Peng, Y.; Cao, C.; Liu, X.; Hou, S.; Chi, J.; Huang, Y.; Li, Q.; and He, Z. 2020. Gaitpart: Temporal part-based model for gait recognition. In *CVPR*, 14225–14233.

Han, X.; Cong, P.; Xu, L.; Wang, J.; Yu, J.; and Ma, Y. 2022. LiCamGait: Gait recognition in the wild by using LiDAR and camera multi-modal visual sensors. *arXiv preprint arXiv:2211.12371*.

Hou, S.; Cao, C.; Liu, X.; and Huang, Y. 2020. Gait lateral network: Learning discriminative and compact representations for gait recognition. In *ECCV*, 382–398. Springer.

Hou, S.; Liu, X.; Cao, C.; and Huang, Y. 2021. Set residual network for silhouette-based gait recognition. *TBIOM*, 3(3): 384–393.

Hou, S.; Liu, X.; Cao, C.; and Huang, Y. 2022. Gait quality aware network: toward the interpretability of silhouette-based gait recognition. *TNNLS*.

Hsu, H.-M.; Wang, Y.; Yang, C.-Y.; Hwang, J.-N.; Thuc, H. L. U.; and Kim, K.-J. 2022. GaitTAKE: Gait recognition by temporal attention and keypoint-guided embedding. In *ICIP*, 2546–2550. IEEE.

Huang, X.; Zhu, D.; Wang, H.; Wang, X.; Yang, B.; He, B.; Liu, W.; and Feng, B. 2021a. Context-sensitive temporal feature learning for gait recognition. In *ICCV*, 12909–12918.

Huang, Z.; Xue, D.; Shen, X.; Tian, X.; Li, H.; Huang, J.; and Hua, X.-S. 2021b. 3D local convolutional neural networks for gait recognition. In *ICCV*, 14920–14929.

Kim, M.; Jain, A. K.; and Liu, X. 2022. Adaface: Quality adaptive margin for face recognition. In *CVPR*, 18750–18759.

Lehmann, R. 2013. 3 σ-rule for outlier detection from the viewpoint of geodetic adjustment. *Journal of Surveying Engineering*, 139(4): 157–165.

Li, P.; Prieto, L.; Mery, D.; and Flynn, P. 2018. Face recognition in low quality images: A survey. *arXiv preprint arXiv:1805.11519*.

Lin, B.; Zhang, S.; and Yu, X. 2021. Gait recognition via effective global-local feature representation and local temporal aggregation. In *ICCV*, 14648–14656.

Loper, M.; Mahmood, N.; Romero, J.; Pons-Moll, G.; and Black, M. J. 2023. SMPL: A skinned multi-person linear model. In *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*, 851–866.

Luo, H.; Gu, Y.; Liao, X.; Lai, S.; and Jiang, W. 2019. Bag of tricks and a strong baseline for deep person re-identification. In *CVPRW*.

Ma, K.; Fu, Y.; Zheng, D.; Cao, C.; Hu, X.; and Huang, Y. 2023. Dynamic aggregated network for gait recognition. In *CVPR*, 22076–22085.

Madhusudana, P. C.; Birkbeck, N.; Wang, Y.; Adsumilli, B.; and Bovik, A. C. 2022. Image quality assessment using contrastive learning. *TIP*, 31: 4149–4161.

Mirelman, A.; Shema, S.; Maidan, I.; and Hausdorff, J. M. 2018. Gait. *Handbook of Clinical Neurology*, 159: 119–134.

Nascimento, R.; Martins, I.; Dutra, T. A.; and Moreira, L. 2023. Computer vision based quality control for additive manufacturing parts. *The International Journal of Advanced Manufacturing Technology*, 124(10): 3241–3256.

Pan, Z.; Zhang, H.; Lei, J.; Fang, Y.; Shao, X.; Ling, N.; and Kwong, S. 2022. DACNN: Blind image quality assessment via a distortion-aware convolutional neural network. *TCSVT*, 32(11): 7518–7531.

Pinyoanuntapong, E.; Ali, A.; Wang, P.; Lee, M.; and Chen, C. 2023. GaitMixer: skeleton-based gait representation learning via wide-spectrum multi-axial mixer. In *ICASSP*, 1–5. IEEE.

Schlett, T.; Rathgeb, C.; Henniger, O.; Galbally, J.; Fierrez, J.; and Busch, C. 2022. Face image quality assessment: A literature survey. *ACM Computing Surveys (CSUR)*, 54(10s): 1–49.

Sepas-Moghaddam, A.; and Etemad, A. 2022. Deep gait recognition: A survey. *TPAMI*, 45(1): 264–284.

Shen, C.; Fan, C.; Wu, W.; Wang, R.; Huang, G. Q.; and Yu, S. 2023. LidarGait: Benchmarking 3D gait recognition with point clouds. In *CVPR*, 1054–1063.

Shen, C.; Yu, S.; Wang, J.; Huang, G. Q.; and Wang, L. 2022. A comprehensive survey on deep gait recognition: algorithms, datasets and challenges. *arXiv preprint arXiv:2206.13732*.

Shiraga, K.; Makihara, Y.; Muramatsu, D.; Echigo, T.; and Yagi, Y. 2016. Geinet: View-invariant gait recognition using a convolutional neural network. In *ICB*, 1–8. IEEE.

Teepe, T.; Gilg, J.; Herzog, F.; Hörmann, S.; and Rigoll, G. 2022. Towards a deeper understanding of skeleton-based gait recognition. In *CVPR*, 1569–1577.

Teepe, T.; Khan, A.; Gilg, J.; Herzog, F.; Hörmann, S.; and Rigoll, G. 2021. GaitGraph: Graph convolutional network for skeleton-based gait recognition. In *ICIP*, 2314–2318. IEEE.

Wan, C.; Wang, L.; and Phoha, V. V. 2018. A survey on gait recognition. *ACM Computing Surveys*, 51(5): 1–35.

Wang, H.; Wang, Y.; Zhou, Z.; Ji, X.; Gong, D.; Zhou, J.; Li, Z.; and Liu, W. 2018. CosFace: Large margin cosine loss for deep face recognition. In *CVPR*, 5265–5274.

Wang, M.; Guo, X.; Lin, B.; Yang, T.; Zhu, Z.; Li, L.; Zhang, S.; and Yu, X. 2023a. DyGait: Exploiting dynamic representations for high-performance gait recognition. *arXiv preprint arXiv:2303.14953*.

Wang, Z.; Hou, S.; Zhang, M.; Liu, X.; Cao, C.; Huang, Y.; and Xu, S. 2023b. LandmarkGait: Intrinsic Human Parsing for Gait Recognition. In *ACM MM*, 2305–2314.

Wu, Q.; Xiao, R.; Xu, K.; Ni, J.; Li, B.; and Xu, Z. 2023. GaitFormer: Revisiting intrinsic periodicity for gait recognition. *arXiv preprint arXiv:2307.13259*.

Yang, S.; Wu, T.; Shi, S.; Lao, S.; Gong, Y.; Cao, M.; Wang, J.; and Yang, Y. 2022. Maniqa: Multi-dimension attention network for no-reference image quality assessment. In *CVPR*, 1191–1200.

Yu, S.; Tan, D.; and Tan, T. 2006. A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In *ICPR*, volume 4, 441–444. IEEE.

Zheng, J.; Liu, X.; Liu, W.; He, L.; Yan, C.; and Mei, T. 2022. Gait recognition in the wild with dense 3d representations and a benchmark. In *CVPR*, 20228–20237.

Zhu, Z.; Guo, X.; Yang, T.; Huang, J.; Deng, J.; Huang, G.; Du, D.; Lu, J.; and Zhou, J. 2021. Gait recognition in the wild: A benchmark. In *ICCV*, 14789–14799.