Free Lunch for Gait Recognition: A Novel Relation Descriptor

Jilong Wang[1,2,4], Saihui Hou[3,4], Yan Huang[2], Chunshui Cao[4], Xu Liu[4], Yongzhen Huang[3,4], Tianzhu Zhang[1], and Liang Wang[2,⋆]

[1] University of Science and Technology of China
[2] Institute of Automation, Chinese Academy of Sciences
[3] Beijing Normal University
[4] WATRIX.AI

Abstract. Gait recognition is to seek correct matches for query individuals by their unique walking patterns. However, current methods focus solely on extracting individual-specific features, overlooking "interpersonal" relationships. In this paper, we propose a novel **Relation Descriptor** that captures not only individual features but also relations between test gaits and pre-selected gait anchors. Specifically, we reinterpret classifier weights as gait anchors and compute similarity scores between test features and these anchors, which re-expresses individual gait features into a similarity relation distribution. In essence, the relation descriptor offers a holistic perspective that leverages the collective knowledge stored within the classifier's weights, emphasizing meaningful patterns and enhancing robustness. Despite its potential, relation descriptor poses dimensionality challenges since its dimension depends on the training set's identity count. To address this, we propose Farthest gait-Anchor Selection to identify the most discriminative gait anchors and an Orthogonal Regularization Loss to increase diversity within gait anchors. Compared to individual-specific features extracted from the backbone, our relation descriptor can boost the performance nearly without any extra costs. We evaluate the effectiveness of our method on the popular GREW, Gait3D, OU-MVLP, CASIA-B, and CCPG, showing that our method consistently outperforms the baselines and achieves state-of-the-art performance.

Keywords: Gait recognition · Relation descriptor · Free lunch

1 Introduction

Gait recognition aims at identifying people at a long distance by their unique walking patterns [33]. As an identification task in vision, the essential goal of it is to learn the distinctive and invariant representations from the physical and behavioral human walking characteristics. With the boom of deep learning, gait recognition has achieved significant progress [6, 11, 12, 34, 39, 47], yielding impressive results on public datasets.

⋆ Corresponding Author

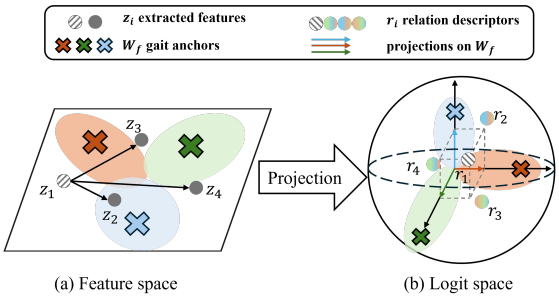


Fig. 1: The comparison of identity-specific embeddings and relation-specific logits. (a) Conventional gait recognition utilizes the extracted features for the final identification. (b) Gait is described by the similarities to fixed semantic directions by projecting gait features on well-trained gait anchor's vectors.

Reappraising an established pipeline of gait recognition [11,18,25,40], it typically involves a *feature extractor* for obtaining unique gait features of individual walking sequences and usually utilizes a *classifier* [26] for accurate identity classification. In common practice, the output embedding of the feature extractor is used for the test, while the classifier is usually discarded since testing and training identities are different. However, the classifier usually occupies the majority of the network's weights, especially when the number of training classes is large, for example, 92.1% and 75.1% of total weights in GaitBase [11] on GREW [53] and OUMVLP [35], respectively. Thus, a natural question arises: *are the well-trained weights in classifiers really useless for inference?*

In this work, we draw wisdom from human beings: **human nature is the ensemble of social relations** [20]. This philosophical point of view provides a new perspective on gait recognition. Gait features are sensitive to many covariant factors such as viewpoints, clothing, and racial differences [17], resulting in challenges of capturing the accurate features under different conditions [5, 39]. Intuitively, providing a relation descriptor within a whole group, such as "tallest person among them", sometimes makes it easier to identify a person than just giving an inherent feature like "7 feet tall". Inspired by it, we assume that gait goes beyond just an aggregation of individual features, and it can also be expressed through the relationships with the gait features of others. As shown in Fig. 1, gait features can be described by the difference/similarity relationships with the several pre-selected people's **G**ait **A**nchors (GAs).

To get such gait descriptors, a set of GAs should be first determined. A good set of GAs should contain various gait patterns. For instance, if we select all GAs with similar body shapes and postures, the relation descriptor loses its discriminative capacity since it fails to reflect distinct features of various gait. In our work, we innovatively find that the weights in the classifier are suitable for severing as GAs by reinterpreting it as *the well-defined gait prototypes of different people in the training set*. Hence, the projection of a gait embedding onto these prototypes can be used as a representation that describes their relationships.

Specifically, the normalized dot product of gait features and GAs generates a distribution of similarities, constructing a new **R**elation **D**escriptor (RD). Essentially, *RD offers a holistic perspective that leverages the collective knowledge stored within the classifier's weights.* In principle, we find RD brings two benefits: 1) emphasizing meaningful features instead of noise and 2) enhancing robustness and generalizability.

In the meanwhile, directly employing RD also poses two challenges, *i.e.*, **dimension expansion** and **GAs overfitting**, as the number of GAs in the classifier depends on the count of training identities. When numerous GAs are selected, such as 20000 and 5153 in GREW [53] and OU-MVLP [35], the dimension of RD can largely surpass that of the original embedding, resulting in increased storage costs and practical challenges in real-world applications. On the other hand, we find that too few identities lead to an overfitting problem that all GAs are highly related, reducing the variety of gait patterns and variations within GAs. Therefore, balancing the quantity of GAs to avoid excessive expansion and removing the correlation between GAs are the key considerations in applying RD.

For the problem of **dimension expansion**, we find that not all identities in the training set are needed since there would be many similar identity prototypes, which results in redundant relationships. Inspired by Archetypal Analysis [46], we assume that the most discriminative combination of GAs in the latent space should be the one with the largest spanning space, *i.e.*, the convex combinations of the archetypes. Accordingly, we introduce a **F**arthest gait-**A**nchor **S**election (FAS) algorithm to select the most discriminative set of GAs. For the problem of **GAs overfitting**, we find the main cause is a target misalignment between cross-entropy loss and GAs. When the number of identities is fewer than the embedding dimension, the network could easily push the logits of the ground truth (GT) class to 1 and the rest to -1, measured by cosine similarity. However, we desire that a sample only related to its own class weight, which means it should be orthogonal to other class weights. Therefore, we propose an **O**rthogonal **R**egularization **L**oss (ORL) for better classifier training, encouraging the cosine similarity of a sample to its own class to be close to 1 and the similarity to other classes to be close to 0 instead of -1. As a result, RD can better reflect the distinct characteristics of different individuals' gaits, even with a small set of GAs.

We rigorously evaluate our proposed approach on GREW [53], Gait3D [52], OU-MVLP [35], CASIA-B [48], and CCPG [22], consistently demonstrating its superiority over conventional baseline methods. To summarize, the contributions of our work can be outlined in three aspects:

(i) We propose a novel descriptor for gait recognition, capturing not only individual features but also relationships among well-trained gait anchors, which enhances recognition performance nearly without extra costs.

(ii) We address the challenges of dimension expansion and GAs overfitting by the Farthest gait-Anchor Selection algorithm and Orthogonal Regularization Loss, improving efficiency and discrimination.

(iii) We evaluate the effectiveness of our proposed method on five popular datasets, and the extensive experimental results demonstrate the superior performance of our approach, *e.g.*, 5.5% and 5.4% absolute improvements on Gait3D and GREW in terms of rank-1 accuracy.

2 Related Work

2.1 Gait Recognition

Gait recognition identifies people by their unique representation of gait characteristics, which is easily affected by many covariant factors. Existing works mainly focus on extracting invariant representation from gait sequences, which can be roughly grouped into model-based [1, 3, 21, 50] and appearance-based [6, 12, 16, 18, 25, 34] categories according to the type of input. Thanks to prior works' contributions, a typical gait recognition pipeline has been established. It primarily consists of four components: 1) a spatial-temporal feature extractor to get individual inherent features (CNN [6, 11, 12, 25, 40], Transformer [10, 44], GCN [13, 36], *etc.*), 2) a temporal pooling module to aggregate sequence's features (MaxPooling [6], MeanPooling [34], GeM [25], *etc.*), 3) a multi-scale module to obtain fine-grained information (attention [8, 9], body parsing [12, 42], *etc.*), 4) and loss functions (Triplet Loss [32], Cross-Entropy Loss [49], *etc.*).

Unlike previous works that emphasize individual-specific gait features, we introduce a novel perspective: a person's gait can be effectively described through the relationships between their gait features and anchor gaits. We also notice a relevant work, GEINet [34], which employs the inner product logits as output without providing specific reasons. Compared to this method, we delve into the discrimination of logits from the perspective of relationships. This analysis not only sheds light on the significance of weights within classifiers but also provides possible solutions to further enhance discrimination. Moreover, we empirically find that the inner product logits are not suitable as test features.

2.2 Open-set Recognition

Open-set recognition (OSR) is a classification task with the additional requirement of rejecting input from unknown classes [30]. In OSR, researchers mainly use the logits from the classifier to measure the distance between "unknown" data and "known" clusters. OpenMax [2] models all known classes by their logits as a single cluster and re-calibrates softmax scores according to the distance between input and other cluster centers. Several following works [14, 27, 29, 31] propose many mechanisms to improve the distance-based measures. Recent works [4, 37] find that a good close-set classifier can directly boost the performance of open-set recognition, leading to a reconsideration of classifiers.

Existing works in OSR also highlight the importance of classifiers. Yet, they primarily differ from our work in two key aspects: (1) The classifier in OSR is needed for "known" class classification, while we drop the concept of classification

and consider training identities as our GAs ; (2) OSR emphasizes the logits of unseen classes should be far from that of "known" classes, while we advocate the discriminative capacity of logits. Overall, our work proposes a brand new perspective for the usage of classifiers.

3 Our Approach

In this work, we propose a new gait descriptor by revisiting the role of classifiers in the testing phase of gait recognition. We innovatively find the well-trained weights in the classifier can be regarded as gait anchors (GAs), and the relationship between the gait features and the set of gait anchors can be used as a discriminative descriptor.

3.1 Pipeline

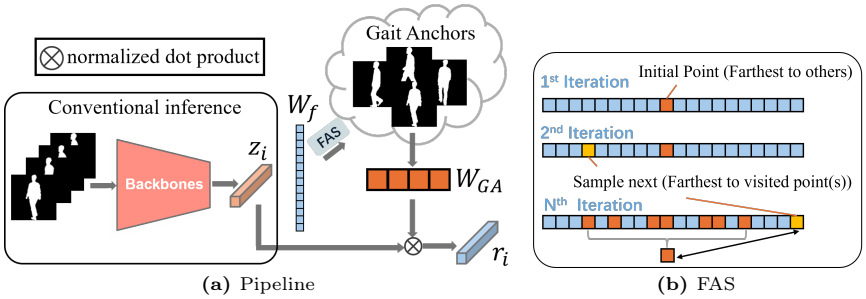
We follow a typical gait recognition procedure [11,25] during the training phase. Given a training set $\mathcal{D} = \{(x_i, y_i)\}$ where x_i denotes a gait sequence and $y_i \in \{0, 1, ..., C\}$ indicates the class label of x_i. The Ω is a feature extractor that embeds the gait sequences into a d-dimensional latent space, where $z_i = \Omega(x_i)$ is the extracted representation. The objective of training Ω on $\mathcal{D}$ is to acquire a discriminative transformation from x_i to $z_i \in \mathbb{R}^d$, ensuring that the distance between z from the same person is closer than those from different people.

A common practice [8,9,11,18,25] uses a combination of triplet loss [32] and cross-entropy (CE) loss [49] to get a discriminative Ω. A BNNeck classifier [26] $f : \mathbb{R}^d \to \mathbb{R}^C$ with a weights matrix $W_f \in \mathbb{R}^{d \times C}$ is introduced during the training stage for the classification task. In our work, we adopt a cosine similarity classifier [15] in BNNeck. The $r_i = f(z_i)$ indicates the normalized dot product results of z_i with W_f. In this context, the logits can be seen as the descriptor of relationships: the larger r_i^j is, the more similar z_i is to the j^{th} weight $W_{f,j}$.

Unlike prior works, we advocate using r_i for the final individual identification, thereby we keep the classifier in the test phase, as in Fig. 2a. To address the mentioned challenges brought by RD, we further adopt Farthest gait-Anchor Selection (FAS) algorithm to select the most discriminative set of GAs and Orthogonal Regularization Loss (ORL) for GAs overfitting problem. Finally, with the help of the well-trained W_f and our proposed methods, relation descriptors r_i that achieve **higher performance** can be obtained **nearly without extra costs**, and their dimensions are **the same or even less** than that of z_i.

3.2 A Novel Relation Descriptor

We now introduce the core idea of our paper: gait goes beyond just an aggregation of individual features, and it can also be expressed through the relationships with the gait features of others. In contrast to conventional methods that use the embedding z_i as the final representation, we find that gait can be expressed by comparing the relationships between the gait features of different individuals,



(a) Pipeline **(b)** FAS

Fig. 2: (a) The overview of our pipeline. The gait anchors are selected from the well-trained classifier and the final representation is projected into a cosine similarity space. (b) Visualization of FAS process.

encompassing aspects of similarity, dissimilarity, common traits, *etc*. For example, given a random person's gait, it can be described as similar to *gait anchor* 1 with 0.2 cosine similarity, 0.7 of GA2, 0.1 of GA3, -0.5 of GA4, and so on. As a result, a descriptor of this person can be formulated as $r = [0.2, 0.7, 0.1, -0.5, ...]$, where each dimension of r denotes the degree of similarity to a gait anchor. We term this novel descriptor **Relation Descriptor** (RD). The Euclidean distance between two RDs can be defined as

$$dist(r_a, r_b) = \sqrt{\Sigma_i^{|GA|}(r_a^i - r_b^i)^2} \tag{1}$$

where $|GA|$ is the number of pre-selected GAs. The potential discriminative capability of RD arises from the observation that gaits from the same person would share similar relationships to GAs, while gaits from different persons tend to have different ones.

Noting that the discriminative capability of the RD heavily relies on the set of GAs, a good GA should be relevant to ID information while remaining unbiased to covariates. The higher the diversity of gait patterns within the GAs group is, the stronger the discrimination capability of the RD is. To get a set of well-defined GAs, we set our sights on the classifier. Particularly, the classifier's weights W_f are optimized using cross-entropy (CE) loss, which ensures that each column vector $W_{f,j}$ encapsulates gait representations uniquely associated with the j^{th} identity and shares less mutual information with other identities in the training set [43, 49]. Given this, the W_f in the classifier is naturally suitable for severing as GAs. Leveraging W_f, we can easily derive our new descriptor RD without any extra costs, formulated as $r_i = \frac{W_f \cdot z_i}{\|W_f\|\|z_i\|}$ where the $\|\cdot\|$ is the L2 norm. The normalized dot product between test gait features and W_f represents similarity. Eq. 1 can be rewritten as

$$d(r_a, r_b) = \sqrt{\Sigma_j^C (\frac{W_{f,j}}{\|W_{f,j}\|} \cdot (\frac{z_a}{\|z_a\|} - \frac{z_b}{\|z_b\|}))^2} \tag{2}$$

It is easily noticed that r_i is exactly the logits used by cosine CE loss [15], which is usually discarded in prior works since it represents the probability distribution of training classification that is inapplicable during testing. Nevertheless, through our novel perspective and experimental results, we argue that RD could also serve as a meaningful and robust representation of gait recognition.

Discussion: Why does RD work? Gait is easily influenced by various covariates. Thus, a gait embedding can be written as $z_i = \hat{z}_i + \epsilon_i$, where $\hat{z}_i$ is the invariant individual feature, and ϵ_i denotes the unexpected bias associated with covariates. An ideal gait recognition model should satisfy the conditions $\epsilon \to 0$ and $dist(z_a, z_p) = 0 < dist(z_a, z_n) - m$, where a, p, n represents the anchor, positive and negative samples, respectively, indicating that z is a consistent representation for the same person while maintaining distinctiveness across different individuals. However, as discussed in [39], it is hard to fulfill these two constraints simply by conventional metric learning, especially on real-world datasets. The embedding z_i inevitably incorporates bias ϵ_i to some extent.

Inspired by image denoising [7] and open-set recognition [28], projecting the original biased embedding z_i onto several ID-relevant bases could suppress the ID-irrelevant bias ϵ_i and retain crucial ID-related information $\hat{z}_i$. This kind of projection decomposes z_i into distinct semantic directions which can be measured by the cosine similarity between z_i and GAs. Fortunately, the well-trained W_f converges toward different ID-related centers, expressed as $W_{f,j} \approx \frac{1}{|\mathcal{D}_j|} \sum_{z_i \in \mathcal{D}_j} (\hat{z}_i + \epsilon_i)$. Since samples of the j^{th} person are usually collected across diverse covariant conditions, the term ϵ_i from different samples would be reduced by the average operator. Thus, the ID center would be less biased, formulated as $W_{f,j} \approx \hat{z}_i + \mathcal{O}(\epsilon)$, where $\mathcal{O}(\epsilon)$ is a minor term of ϵ. Consequently, the W_f is naturally more independent of covariates.

In summary, RD can be seen as the projection of z onto different GAs, resulting in reduced susceptibility to the influence of covariates.

3.3 Challenge: Dimension Expansion

As we discussed above, the GAs can be seen as a set of semantic bases in the latent space, but, the number of bases relies on the number of training identities. When training on a large-scale dataset, the dimension of W_f is inevitably expanding, incurring augmented storage costs and redundant information. Therefore, the combination of GAs should be carefully selected.

What is a good combination of gait anchors? The latent space is a manifold determined by the training data, and the class weights can be seen as bases that span the manifold. Intuitively, in order to maintain the discriminative ability, the manifold spanned by the combination of selected GAs needs to be as consistent as possible with the original manifold. Additionally, when the number of identities is larger than the dimension of z_i, $C > d$, there are many linearly related bases in the classifier that can be removed.

Based on the above discussion, we assume that the most discriminative combination of GAs should be the one with the largest spanning space, *i.e.*, the convex combinations of the GAs.

Algorithm 1 Pseudo-code of FAS in a PyTorch-like style.

```
# cos_sim(): matrix-wise cosine similarity
dist = abs(cos_sim(W_f, W_f.t()))
# the initial farthest weight (dist -> 0) from all others
farthest = dist.sum(-1).argmin()
W_fs[0] = W[farthest]
# remove the selected basis from W
W_f.remove(farthest)

for i in range(1, N):
    dist = abs(cos_sim(W_fs[i-1], W_f.t()))
    farthest = dist.argmin()
    W_fs[i] = W[farthest]
    W_f.remove(farthest)
```

How to select good gait anchors. Selecting the convex combinations of GAs from W_f in the classifier is an NP-hard problem, which is generally difficult to solve in polynomial time. Hence, we propose a heuristic method, called Farthest gait-Anchor Selection (FAS), to solve this problem.

Given weights in the classifier $\{W_{f,1}, W_{f,2}, ..., W_{f,C}\}$, we use iterative FAS to choose a combination of N weights $\{W_{f,s1}, W_{f,s2}, ..., W_{f,sN}\}$ as GAs, where the $W_{f,sj}$ is the farthest weights of $W_{f,sj-1}$, measured by absolute cosine similarity (0 is the farthest). The final selected weights are denoted by $W_{f,s} \in \mathbb{R}^{d\times N}$. FAS is a greedy algorithm with a time complexity of $\mathcal{O}(NlogC)$ described in Algorithm 1 and a visualization of this process is shown in Fig. 2b. Note that FAS **only runs once** after the training is completed, and $W_{f,s}$ are fixed for all test samples during the inference phase.

The internal mechanism of FAS is based on the assumption that the farthest weight from the selected weights contains the most different semantic information, thus, it is good for increasing the diversity of the combination gait anchors. Compared to the original W_f, the $W_{f,s}$ collects the discriminative gait anchors and removes redundant ones.

How to reduce the final dimension. After filtering out some useless weights by FAS, the number of gait anchors may be still larger than the original embeddings' dimension. Luckily, the dimension of $W_{f,s}$ can be reduced to d without information loss by Singular Value Decomposition (SVD) when $d < |GA|$. *Note that directly applying SVD to reduce the dimension of W_f cannot bring discriminative improvements in GAs*, since SVD is an identical transformation of W_f. As a result, SVD here is only used to transform the dimension of $W_{f,s}$ from $\mathbb{R}^{d\times N}$ to $\mathbb{R}^{d\times d}$ when $d < N$, ensuring r_i has the same dimension as z_i.

3.4 Challenge: GAs overfitting

As previously discussed, the discrimination capability of RD relies on the diversity of gait patterns within the GA group, necessitating that each GA holds different semantic directions in the latent space. Consequently, the desirable GAs should share as little as correlated information, mathematically, requiring orthog-

onality in the latent space. However, this requirement encounters challenges when the number of training identities is limited, particularly when $C < d$. In such instances, the classifier is prone to overfitting, leading to high inter-relatedness among GAs.

An analysis of the CE loss reveals a misalignment with the ideal characteristics of GAs. During the training stage, the optimization target of CE is

$$\mathcal{L}_{ce} = log(1 + \sum_{y' \neq y_i} exp((r^{y'} - r^{y_i})/T)) \tag{3}$$

where T is the temperature and r^{y_i} indicates the ground truth y_i^{th} index of logits while $r^{y'}$ are the rest of logits. Notably, L_{ce} encourages that r^{y_i} is close to 1 while $r^{y'}$ is close to -1. However, as previously mentioned, the desired $r^{y'}$ should be 0 (orthogonal) instead of -1 (negative correlation), where z_i only holds a positive correlation to ground truth class weight W_{f,y_i}.

To tackle this limitation, we propose an orthogonal regularization loss (ORL) to minimize the correlation between different identity weights, increasing the diversity among GAs, which can be formulated as

$$\mathcal{L}_{ORL} = \frac{1}{|\mathcal{B}|} \sum_{i=1}^{\mathcal{B}} (1 - r_i^{y_i} + \frac{1}{C-1} \sum_{y' \neq y_i} |r_i^{y'}|) \tag{4}$$

where $\mathcal{B}$ is the mini-batch, C is the total number of training identities, and $|r_i^{y'}|$ denotes the absolute value of $r_i^{y'}$.

By enhancing the orthogonality between gait anchors and samples, the RD can better reflect the distinct characteristics of different individuals' gaits even with a small number of anchors. A more carefully designed method is sure to further improve the performance, meriting more exploration in the future. Note that ORL is only used in small datasets with only a few identities.

3.5 Optimization and Inference

Our method can be built on top of off-the-shelf methods without extra parameters. The entire objective function can be formulated as

$$\mathcal{L} = \mathcal{L}_{tri} + \mathcal{L}_{ce} + \lambda \mathcal{L}_{ORL} \tag{5}$$

where the λ is a hyper-parameter.

4 Experiments

4.1 Datasets

GREW [53]. GREW is the largest outdoor dataset, containing 26,345 subjects with 128,671 sequences captured from 882 cameras. According to its official

Table 1: Implementation details and Hyper-parameters.

Dataset	Batch Size	Optimizer	Initial Lr.	Lr. Drop	Total Steps N	FAS λ	of ORL
GREW	(32, 4)	SGD	0.1	(80k, 120k, 150k)	180k	8192	-
Gait3D	(32, 4)			(20k, 40k, 50k)	60k	1024	-
OU-MVLP	(32, 8)			(60k, 80k, 100k)	120k	2048	-
CASIA-B	(8, 16)			(20k, 40k, 50k)	60k	74	1.0
CCPG	(8, 16)			(20k, 40k, 50k)	60k	100	0.1

partition, GREW is divided into three subsets, *i.e.*, the training set with 20k subjects, the validation set with 345 subjects, and the test set with 6k subjects.
Gait3D [52]. Gait3D is an in-the-wild dataset with 4k subjects and over 25k sequences captured from 39 cameras. Gait3D provides an official protocol where 3k subjects are used for training while the remaining 1k subjects are for test.
OU-MVLP [35]. The OU-MVLP is the largest indoor gait dataset under a fully controlled environment. It includes 10,307 subjects under normal walking conditions and 14 views. We adopt the widely-used protocol that 5153 subjects are used for training, and the rest are taken for the test.
CASIA-B [48]. It is one of the most popular gait datasets, which contains 124 subjects from 11 view angles and 3 walking conditions: normal walking (NM), carrying bags (BG), and wearing a coat or jacket (CL). Our work follows the popular partition [6] where the first 74 subjects are used for the training stage and the remaining 50 subjects are reserved for the test.
CCPG [22]. It is a clothing-changing dataset that provides 200 identities and over 16K sequences. Each identity has seven different cloth-changing statuses. CCPG-G is a subset of it, which provides off-the-shelf silhouettes for the gait recognition task. According to its official protocol, we use the first 100 identities for training and the rest for the test.

4.2 Implementation Details

We utilize GaitBase [11] as our main baseline on all datasets. The new version of GaitGL [24], with a backbone of (64,128,128) channels, is also used on OU-MVLP, CASIA-B, and CCPG for a comprehensive comparison. We use a cosine similarity BNNeck classifier for all models, and the temperature in CE loss is set to 16. We reproduce and train all methods by OpenGait [11], following the implementation settings presented in Table 1.

4.3 Performance Comparison

GREW. We compare the performance of the proposed method with several gait recognition methods on GREW dataset and show experimental results in Table 2a. GREW is collected under an unconstrained condition, and it contains lots of unpredictable external covariates, such as occlusion and bad segmentation. As a result, gait sequences in the test set may be encoded by some unseen covariates that further produce meaningless gait representations. From Table 2a,

Table 2: (a) Rank-1 accuracy (%), Rank-5 accuracy (%), Rank-10 accuracy (%), and Rank-20 accuracy (%) on GREW dataset.(b) Rank-1 accuracy(%), mAP(%), and mINP(%) comparison on Gait3D. The **bold** number denotes the best performances.

Methods	Rank-1	Rank-5	Rank-10	Rank-20
PoseGait [23]	0.2	1.0	2.2	4.3
GaitGraph [36]	1.3	3.5	5.1	7.5
GEINet [34]	6.8	13.4	17.0	21.0
TS-CNN [45]	13.6	24.6	30.2	37.0
GaitSet [6]	46.3	63.6	70.3	76.8
GaitPart [12]	44.0	60.7	67.3	73.5
GaitGL [25]	47.3	63.6	69.3	74.2
MGN [38]	44.5	61.3	67.7	72.7
CSTL [19]	50.6	65.9	71.9	76.9
MTSGait [51]	55.3	71.3	76.9	81.6
GaitBase [11]	60.1	75.7	80.5	84.4
↪ ***w.*** ours	**65.5**	**78.7**	**83.3**	**86.3**

(a) GREW

Methods	Gait3D		
	Rank-1	mAP	mINP
PoseGait [23]	0.2	0.5	0.3
GaitGraph [36]	6.3	5.2	2.4
GaitSet [6]	36.7	30.0	17.3
GaitPart [12]	28.2	21.6	12.4
GLN [18]	31.4	24.7	13.6
GaitGL [25]	29.7	22.3	13.3
CSTL [19]	11.7	5.6	2.6
SMPLGait [52]	46.3	37.2	22.2
GaitBase [11]	64.6	55.2	36.2
↪ ***w.*** ours	**70.1**	**61.9**	**41.3**

(b) Gait3D

we can see the gait recognition methods that perform well on indoor datasets meet a large performance degradation. It shows the gait representations encoded by only individual features are not robust enough. By replacing our relation descriptor incorporating FAS and SVD, our method elevates the accuracy of GaitBase by 5.4% on Rank-1 accuracy. It is worth noting that our method adds no extra parameters and the dimension of the final gait representation is the same as the original GaitBase. The experimental results indicate that the relation descriptor is more discriminative and robust in real-world scenarios.

Gait3D.Gait3D is also an unconstrained dataset. The comparison of prevailing competing methods is illustrated in Table 2b, demonstrating that our proposed method exhibits superior performance compared to previous methods by a considerable margin. Our method measures the relationship between test gait and well-defined gait anchors, which is less affected by unseen covariates. As a result, our method outperforms all prevailing methods and boosts the performance of GaitBase by 5.4%, 6.7%, and 5.8% on Rank-1, mAP, and mINP, respectively.

OU-MVLP. Since OU-MVLP is collected in a fully-constrained laboratory environment, its training and testing sets possess nearly identical covariates. The effects of covariates can be well eliminated during the training stage. Hence, directly using individual gait features as representation can achieve promising accuracy. However, as shown in Table 3, we find our method can still boost the baseline performance by using RD with the same dimension as the corresponding embedding on this dataset.

CASIA-B and CCPG-G. CASIA-B and CCPG-G are also collected in a fully-constrained environment, containing viewpoints and clothing changes. It is worth noting that there are only 74 and 100 identities in the training set in CASIA-B and CCPG-G, which means the initial dimensions of RD are 74 and 100 on these two datasets, respectively. As shown in Table 3, even though the dimension of RD is fewer than the original embeddings, our method still achieves higher

Table 3: Averaged rank-1 accuracy (%) on CASIA-B and OU-MVLP datasets. Rank-1 accuracy (%) and mAP on CCPG-G dataset. The '*' denotes that the results are based on our strict reproduction by OpenGait.

Method	CASIA-B			CCPG-G (Rank-1 | mAP)						OU-MVLP
	NM	BG	CL	CL		UP		DN		mean
GEINet [34]	48.1	35.7	23.5	-		-		-		42.5
GaitSet [6]	95.8	90.0	75.4	77.7	46.5	83.5	59.6	83.2	60.1	87.1
GaitPart [12]	96.1	90.7	78.7	77.8	45.5	84.5	63.1	83.3	60.1	88.7
GLN [18]	96.9	94.0	77.5	-		-		-		89.2
CSTL [19]	98.0	**95.4**	87.0	-		-		-		90.2
GaitGL* [24]	97.7	94.7	86.0	81.9	50.5	91.2	71.0	86.4	67.0	89.7
↪ ***w.*** ours	97.5	95.0	**87.8**	82.4	51.8	91.4	72.3	87.2	69.5	90.5
GaitBase* [11]	97.6	94.0	77.4	91.8	64.6	95.3	78.3	94.7	79.5	90.8
↪ ***w.*** ours	**98.1**	94.1	77.9	**92.3**	**67.3**	**95.5**	**81.6**	**95.3**	**81.6**	**91.3**

accuracy. Our method improves the average mAP of Gaitbase by 2.7% on CCPG and 0.6% on CASIA-B. Since GaitGL performs better than GaitBase on CASIA-B, we also adapt our method to GaitGL, and the results on both datasets again verify the effectiveness of our method. The experiments have demonstrated that relation descriptors exhibit discriminative capability comparable to or higher than individual gait features.

4.4 Ablation Study

In this subsection, we provide the ablation study of each component in our method on CASIA-B and Gait3D with GaitGL and GaitBase.

Analysis of RD, FAS, and ORL. The ablation results are illustrated in Table 4a. Note that FAS is only employed for datasets where the number of identities C in the training set exceeds the output dimension d. In contrast, ORL is only applied on datasets where $C < d$. Here is the analysis: 1) from #2, RD improves the baseline by 2.2% on Gait3D but degrades the performance on CASIA-B by 1.4% due to fewer GAs; 2) Comparing #4 with #2, adding ORL to generate approximately orthonormal GAs brings improvement on CASIA-B, which shows the effectiveness of ORL; 3) Comparing #4 with #2, the selected combination of GAs by FAS can further improve the recognition accuracy. Overall, the ablation results verify the effectiveness of our proposed methods.

Analysis of model-agnostic results. As shown in Table 4b, we conduct experiments on Gait3D with four popular methods, including GaitPart, GPGait, QAGait, and DyGait. Since GaitPart doesn't use CE loss, we train an extra cosine similarity BNNeck classifier for it. The results show that the proposed method can improve performance regardless of the baseline backbones.

Analysis of weight λ of ORL. Tab. 4c shows that the mean accuracy on CASIA-B varies with different λ values. The optimal λ lies between 0.1 and 2.0.

Analysis of the generalizable ability on cross-domain testing. As shown in Fig. 3a, we compare the cross-domain recognition performance between RD r_i

Table 4: (a) Ablation study on relation descriptors (RD), Farthest gait-Anchor Selection (FAS), and Orthogonal Regularization Loss (ORL) on CASIA-B with GaitGL and Gait3D with GaitBase. (b) Model-agnostic results on Gait3D dataset. (c) Sensitivity analysis of weight λ of ORL on CASIA-B

	RD	FAS	ORL	CASIA-B	Gait3D
				Mean Acc. ↑	Rank-1 Acc. ↑
#1				92.8	64.6
#2	✓			91.4$^{\downarrow-1.4}$	68.8$^{\uparrow+2.2}$
#3	✓	✓		*N/A*	**70.1**$^{\uparrow+5.5}$
#4	✓		✓	**93.4**$^{\uparrow+0.6}$	*N/A*

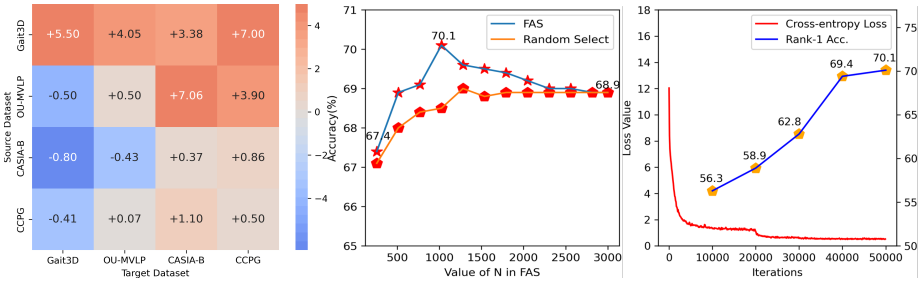
(a)

Methods	GaitPart	GPGait [13]	QAGait [41]	DyGait [40]
Embedding z	28.2	22.4	67.0	68.3
↪ $\boldsymbol{w}$. ours	38.0$^{\uparrow+9.8}$	27.9$^{\uparrow+5.5}$	70.8$^{\uparrow+3.7}$	71.4$^{\uparrow+3.1}$

(b)

λ	0.0	0.1	0.5	1.0	2.0	5.0
Mean Acc.	91.4	92.1	93.1	**93.4**	92.9	90.5

(c)



(a) **(b)** **(c)**

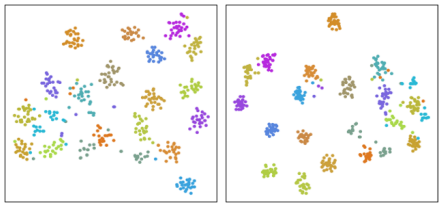
Fig. 3: (a) The gain of recognition accuracy (+/−) brought by RD r for cross-domain testing compared to using embedding z among Gait3D, OU-MVLP, CASIA-B, and CCPG. (b) Ablation study on the number of selected gait anchors in FAS on Gait3D. (c) The cross-entropy loss and accuracy curves of GaitBase on Gait3D.

and embedding z_i of GaitBase. In most cases, RD gains a better rank-1 accuracy, such as 7.0% improvements on Gait3D to CCPG compared to embedding. This experiment demonstrates that RD potentially alleviate the domain gap.

Analysis of the value N in FAS. As mentioned above, not all weights in the classifier contribute to better recognition performance. Fig. 3b shows that the performance of relation descriptors relies on the pre-selected gait anchors, and selecting all of the weights doesn't necessarily lead to performance improvement, where the curve first rises and then falls as N is increased. Moreover, FAS can bring consistent performance improvement compared to random selection.

Analysis of the relationship between classifier convergence and accuracy. RD relies on well-defined classifier weights, assuming each weight represents a typical gait representation. CE loss requires sufficient separation among classifier weights. As shown in Fig. 3c, classifier weights with better convergence could represent more distinct gait prototypes, leading to superior results.

Analysis of the impact of cross-entropy loss and the importance of cosine similarity As shown in Table 5, our method shows the effectiveness on both vanilla CE Loss (single FC Layer) and BNNecks. The results further demon-



(a) Embedding **(b)** RD

Fig. 4: T-SNE visualization on Gait3D test set with randomly selected 20 classes. One colour denotes a class.

Table 5: Ablation study of CE loss types and comparison of inner product (in GEINet [34]) and cosine similarity logits. The numbers within parentheses indicate the accuracy using z_i.

	$\mathcal{L}_{ce}^{vanilla}$	$\mathcal{L}_{ce}^{BN}$	Gait3D	CASIA-B
			Rank-1 Acc.	Mean Acc.
Inner Prod.	✓		59.7 (**61.8**)	89.9 (**92.8**)
		✓	**65.0** (64.7)	91.3 (**92.9**)
Cos Sim.	✓		**68.5** (63.4)	**92.1** (91.7)
		✓	**70.1** (64.6)	**93.4** (92.8)

strate that cosine similarity is preferred as the relation descriptor, attributed to its robustness and clear interpretability. Besides, since the inner product is an unbounded operator, the value of logits is susceptible to magnitude changes, making the Euclidean distance between them less meaningful in such cases.

Visualization We visualize the feature distribution of models' original embedding and RD on Gait3D test set with 20 randomly selected classes. By comparing Figs. 4a and 4b, it shows that the intra-class variation is further reduced and the inter-class distance is hence enlarged by our relation descriptor.

5 Limitations and Conclusion

GA relies on labeled identities in the training set, limiting its use in unsupervised scenarios. Besides, when covariates are well reduced during training (e.g., OUMVLP), RD and embedding perform comparably. Future work may involve integrating learnable GA generation techniques to address these issues.

In conclusion, we provide a new perspective that gait can be expressed by a relation descriptor by projecting to ID-related GAs, which offers a way to diminish the bias in original embeddings. Further, we revisit the role of the well-trained weights in the classifier, arguing that they are exactly as suitable GAs. Based on this finding, we propose a novel relation descriptor serving as a more discriminative representation of gait recognition. Besides, FAS and ORL are carefully designed to solve dimensionality challenges brought by RD and further boost recognition performance. Overall, we hope these new insights could prompt further research in the gait community.

Acknowledgment

This work was jointly supported by National Key R&D Program of China (2022ZD0117900), National Natural Science Foundation of China (62236010, 62322607, 62276261, 62276025 and 62206022), Beijing Municipal Science & Technology Commission (Z231100007423015) and Shenzhen Technology Plan Program (KQTD20170331093217368).

References

1. Ariyanto, G., Nixon, M.S.: Model-based 3d gait biometrics. In: 2011 International Joint Conference on Biometrics (IJCB). pp. 1–7 (2011)
2. Bendale, A., Boult, T.E.: Towards open set deep networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1563–1572 (2016)
3. Bodor, R., Drenner, A., Fehr, D., Masoud, O., Papanikolopoulos, N.: View-independent human motion classification using image-based reconstruction. Image Vision Comput. **27**(8), 1194–1206 (jul 2009)
4. Cen, J., Luan, D., Zhang, S., Pei, Y., Zhang, Y., Zhao, D., Shen, S., Chen, Q.: The devil is in the wrongly-classified samples: Towards unified open-set recognition. arXiv preprint arXiv:2302.04002 (2023)
5. Chai, T., Mei, X., Li, A., Wang, Y.: Silhouette-based view-embeddings for gait recognition under multiple views. In: 2021 IEEE international conference on image processing (ICIP). pp. 2319–2323. IEEE (2021)
6. 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**(7), 3467–3478 (2021)
7. Cheng, S., Wang, Y., Huang, H., Liu, D., Fan, H., Liu, S.: Nbnet: Noise basis learning for image denoising with subspace projection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 4896–4906 (2021)
8. Dou, H., Zhang, P., Su, W., Yu, Y., Li, X.: Metagait: Learning to learn an omni sample adaptive representation for gait recognition. In: European Conference on Computer Vision. pp. 357–374. Springer (2022)
9. Dou, H., Zhang, P., Su, W., Yu, Y., Lin, Y., Li, X.: Gaitgci: Generative counterfactual intervention for gait recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5578–5588 (2023)
10. Fan, C., Hou, S., Huang, Y., Yu, S.: Exploring deep models for practical gait recognition. arXiv preprint arXiv:2303.03301 (2023)
11. Fan, C., Liang, J., Shen, C., Hou, S., Huang, Y., Yu, S.: Opengait: Revisiting gait recognition towards better practicality. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 9707–9716 (June 2023)
12. Fan, C., Peng, Y., Cao, C., Liu, X., Hou, S., Chi, J., Huang, Y., Li, Q., He, Z.: GaitPart: Temporal part-based model for gait recognition. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 14213–14221 (2020)
13. Fu, Y., Meng, S., Hou, S., Hu, X., Huang, Y.: Gpgait: Generalized pose-based gait recognition. arXiv preprint arXiv:2303.05234 (2023)
14. Ge, Z., Demyanov, S., Chen, Z., Garnavi, R.: Generative openmax for multi-class open set classification. arXiv preprint arXiv:1707.07418 (2017)
15. Gidaris, S., Komodakis, N.: Dynamic few-shot visual learning without forgetting. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4367–4375 (2018)
16. Han, J., Bhanu, B.: Individual recognition using gait energy image. IEEE Transactions on Pattern Analysis and Machine Intelligence **28**(2), 316–322 (2006)
17. Hill, C.N., Reed, W., Schmitt, D., Sands, L.P., Queen, R.M.: Racial differences in gait mechanics. Journal of biomechanics **112**, 110070 (2020)
18. Hou, S., Cao, C., Liu, X., Huang, Y.: Gait lateral network: Learning discriminative and compact representations for gait recognition. In: Computer Vision - ECCV

2020: 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part IX. pp. 382–398. Springer-Verlag, Berlin, Heidelberg (2020)
19. Huang, X., Zhu, D., Wang, H., Wang, X., Yang, B., He, B., Liu, W., Feng, B.: Context-sensitive temporal feature learning for gait recognition. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 12909–12918 (2021)
20. Jaggar, A.M.: Feminist politics and human nature. Rowman & Littlefield (1983)
21. 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. pp. 1058–1064 (2009)
22. Li, W., Hou, S., Zhang, C., Cao, C., Liu, X., Huang, Y., Zhao, Y.: An in-depth exploration of person re-identification and gait recognition in cloth-changing conditions. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 13824–13833 (2023)
23. 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 (2020)
24. Lin, B., Zhang, S., Wang, M., Li, L., Yu, X.: Gaitgl: Learning discriminative global-local feature representations for gait recognition. arXiv preprint arXiv:2208.01380 (2022)
25. Lin, B., Zhang, S., Yu, X.: Gait recognition via effective global-local feature representation and local temporal aggregation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 14648–14656 (October 2021)
26. Luo, H., Jiang, W., Gu, Y., Liu, F., Liao, X., Lai, S., Gu, J.: A strong baseline and batch normalization neck for deep person re-identification. IEEE Transactions on Multimedia **22**(10), 2597–2609 (2019)
27. Miller, D., Sunderhauf, N., Milford, M., Dayoub, F.: Class anchor clustering: A loss for distance-based open set recognition. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 3570–3578 (2021)
28. Miller, D., Sunderhauf, N., Milford, M., Dayoub, F.: Class anchor clustering: A loss for distance-based open set recognition. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 3570–3578 (2021)
29. Neal, L., Olson, M., Fern, X., Wong, W.K., Li, F.: Open set learning with counterfactual images. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 613–628 (2018)
30. Scheirer, W.J., de Rezende Rocha, A., Sapkota, A., Boult, T.E.: Toward open set recognition. IEEE transactions on pattern analysis and machine intelligence **35**(7), 1757–1772 (2012)
31. Schlachter, P., Liao, Y., Yang, B.: Open-set recognition using intra-class splitting. In: 2019 27th European signal processing conference (EUSIPCO). pp. 1–5. IEEE (2019)
32. Schroff, F., Kalenichenko, D., Philbin, J.: Facenet: A unified embedding for face recognition and clustering. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 815–823 (2015)
33. Shen, C., Yu, S., Wang, J., Huang, G.Q., Wang, L.: A comprehensive survey on deep gait recognition: algorithms, datasets and challenges. arXiv preprint arXiv:2206.13732 (2022)
34. 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). pp. 1–8 (2016)

35. 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**(1) (2018)
36. Teepe, T., Khan, A., Gilg, J., Herzog, F., Hörmann, S., Rigoll, G.: Gaitgraph: Graph convolutional network for skeleton-based gait recognition. In: 2021 IEEE International Conference on Image Processing (ICIP). pp. 2314–2318. IEEE (2021)
37. Vaze, S., Han, K., Vedaldi, A., Zisserman, A.: Open-set recognition: A good closed-set classifier is all you need? arXiv preprint arXiv:2110.06207 (2021)
38. Wang, G., Yuan, Y., Chen, X., Li, J., Zhou, X.: Learning discriminative features with multiple granularities for person re-identification. In: Proceedings of the 26th ACM international conference on Multimedia. pp. 274–282 (2018)
39. Wang, J., Hou, S., Huang, Y., Cao, C., Liu, X., Huang, Y., Wang, L.: Causal intervention for sparse-view gait recognition. In: Proceedings of the 31st ACM International Conference on Multimedia. pp. 77–85 (2023)
40. Wang, M., Guo, X., Lin, B., Yang, T., Zhu, Z., Li, L., Zhang, S., Yu, X.: Dygait: Exploiting dynamic representations for high-performance gait recognition. arXiv preprint arXiv:2303.14953 (2023)
41. Wang, Z., Hou, S., Zhang, M., Liu, X., Cao, C., Huang, Y., Li, P., Xu, S.: Qagait: Revisit gait recognition from a quality perspective. In: Proceedings of the AAAI Conference on Artificial Intelligence. pp. 5785–5793 (2024)
42. Wang, Z., Hou, S., Zhang, M., Liu, X., Cao, C., Huang, Y., Xu, S.: Landmarkgait: Intrinsic human parsing for gait recognition. In: Proceedings of the 31st ACM International Conference on Multimedia. pp. 2305–2314 (2023)
43. Wen, Y., Zhang, K., Li, Z., Qiao, Y.: A discriminative feature learning approach for deep face recognition. In: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part VII 14. pp. 499–515. Springer (2016)
44. Wu, Q., Xiao, R., Xu, K., Ni, J., Li, B., Xu, Z.: Gaitformer: Revisiting intrinsic periodicity for gait recognition. arXiv preprint arXiv:2307.13259 (2023)
45. 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**(2), 209–226 (2016)
46. Xiong, Y., Liu, W., Zhao, D., Tang, X.: Face recognition via archetype hull ranking. In: Proceedings of the IEEE international conference on computer vision. pp. 585–592 (2013)
47. Yu, S., Chen, H., Reyes, E.B.G., Poh, N.: Gaitgan: Invariant gait feature extraction using generative adversarial networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). pp. 532–539 (2017)
48. Yu, S., Tan, D., Tan, T.: A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In: 18th international conference on pattern recognition (ICPR'06). vol. 4, pp. 441–444. IEEE (2006)
49. Zhang, Z., Sabuncu, M.: Generalized cross entropy loss for training deep neural networks with noisy labels. Advances in neural information processing systems **31** (2018)
50. Zhao, G., Liu, G., Li, H., Pietikainen, M.: 3d gait recognition using multiple cameras. In: 7th International Conference on Automatic Face and Gesture Recognition (FGR06). pp. 529–534. IEEE (2006)
51. Zheng, J., Liu, X., Gu, X., Sun, Y., Gan, C., Zhang, J., Liu, W., Yan, C.: Gait recognition in the wild with multi-hop temporal switch. In: Proceedings of the 30th ACM International Conference on Multimedia. pp. 6136–6145 (2022)

52. Zheng, J., Liu, X., Liu, W., He, L., Yan, C., Mei, T.: Gait recognition in the wild with dense 3d representations and a benchmark. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 20228–20237 (2022)
53. Zhu, Z., Guo, X., Yang, T., Huang, J., Deng, J., Huang, G., Du, D., Lu, J., Zhou, J.: Gait recognition in the wild: A benchmark. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 14789–14799 (2021)