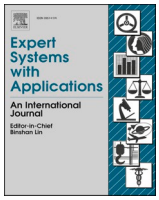
Expert Systems With Applications 244 (2024) 122975

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa



An innovative unsupervised gait recognition based tracking system for safeguarding large-scale nature reserves in complex terrain

Chichun Zhou [a,*,1], Xiaolin Guan [a,1], Zhuohang Yu [a], Yao Shen [b,*], Zhenyu Zhang [a,*], Junjie Gu [c]

[a] School of Engineering, Dali University, Air-Space-Ground Integrated Intelligence and Big Data Application Engineering Research Center of Yunnan Provincial Department of Education, Yunnan 671003, China
[b] School of Criminal Investigation, People's Public Security University of China, Beijing 100038, China
[c] Department of Mechanical and Aerospace Engineering, Carleton University, 1125 Colonel By Drive, Ottawa, Ontario K1S 5B6, Canada

ARTICLE INFO

Keywords:
Unsupervised Gait Recognition
Safeguarding Large-scale Nature Reserves in Complex Terrain
Unsupervised Gait Recognition based Tracking System

ABSTRACT

Abnormal human activities play a significant role in triggering emergencies within vast nature reserves. The vast area, complex terrain, and insufficient electricity and high-bandwidth network infrastructure present significant challenges in effectively supervising nature reserves. Fortunately, intricate terrains often boast restricted access points, typically confined to just a few narrow pathways and the gait recognition technique utilizes only a small amount of binary-processed low-quality gait data and seamlessly integrates with low-resolution and low-power-consumption cameras making it particularly suitable for human activities supervision in nature reserves. However, extensive existing supervised along with a limited number of unsupervised methods are unable to be implemented in real-world application due to the reliance on the pre-labeled training set and the insufficient retrieval accuracies. Here, we present an electronic tracking system for safeguarding large-scale nature reserves in complex terrain based on the unsupervised gait recognition technique for the first time. 1) The proposed method doesn't require any known matching relationships in the training set. 2) It consistently achieves 100% top-1 retrieval accuracies, with a distinct gap between the distances of top-1 and top-2 retrievals. This distinction allows us to detect abnormal behaviors, such as individuals who enter without exiting, exit without entering, or venture into restricted areas. It effectively mitigates the impact of human activities on the protected area at low cost offering an application case of gait recognition technology (GRT) in the field of nature conservation.

1. Introduction

Abnormal human activities play a significant role in triggering emergencies, such as fires, illegal hunting, incidents involving missing persons, etc., within expansive nature reserves characterized by intricate terrains. The management, emergency response, and rescue operations of these nature reserves demand substantial annual investments. For instance, in the past decade, a conservative estimate suggests that human activities of non-regular visitors were responsible for over 85 % of fires in Cangshan Mountain, a famous mountain in China. These visitors, who often access the area through undeveloped tourist route, pose a greater risk of triggering emergencies, especially fires. The annual cost associated solely with fire prevention and emergency response have surpassed 10 million RMB. Hence, there is an urgent need for cost-effective and efficient solutions to effectively monitor and supervise human activities, especially those of non-regular visitors, within nature reserves.

However, the extensive expanse and intricate topography of nature reserves pose challenges to traditional manual inspection methods when it comes to monitoring and managing individuals entering these areas. Although advanced biometric recognition technology utilizing artificial intelligence, such as the Eye-in-the-Sky system (Singh, Patil, & Omkar, 2018), has matured and offers real-time tracking, observation, and recording of targeted activities and behaviors for comprehensive surveillance, the absence of electricity and high-bandwidth network infrastructure within the reserves hampers the implementation of existing such techniques. For instance, they rely on high-definition and high-power-consumption equipment and high-bandwidth networks,

* Corresponding authors.
E-mail addresses: zhouchichun@dali.edu.cn (C. Zhou), guanxiaolin@stu.dali.edu.cn (X. Guan), yuzhuohang@stu.dali.edu.cn (Z. Yu), shenyaophysics@hotmail.com (Y. Shen), zhangzhenyu@dali.edu.cn (Z. Zhang), junjie.gu@carleton.ca (J. Gu).
[1] These authors contribute equivalently.

https://doi.org/10.1016/j.eswa.2023.122975
Received 7 October 2023; Received in revised form 13 December 2023; Accepted 14 December 2023
Available online 16 December 2023

resulting in substantial costs for infrastructure development. Currently, there are some, but relatively fewer efforts, in search of more improved solutions (Isabelle & Westerlund, 2022; Liu et al., 2022; Pickering, Rossi, Hernando, & Barros, 2018). For instance, Liu et al. (2022) used GPS trajectory data to analyze patterns of tourist transfer behavior and spatio-temporal movement behavior. Nevertheless, the vast area, complex terrain, and insufficient electricity and high-bandwidth network infrastructure present significant hurdles in effectively supervising nature reserves.

Gait recognition is a technology that utilizes human walking characteristics to identify individual identities. Its advantages, including non-intrusiveness, long-range capability, and resistance to forgery, make it highly applicable in diverse areas such as security monitoring, medical diagnosis, and intelligent interaction (Sepas-Moghaddam & Etemad, 2022; Singh, Jain, Arora, & Singh, 2018). The academic community has shown substantial interest in this field, introducing various gait datasets and proposing different algorithms (Wan, Wang, & Phoha, 2018; Wu, & Xu, 2019; Zhang et al., 2020; Su, Zhao, & Li, 2020; Arshad et al., 2020; Chao, Wang, He, Zhang, & Feng, 2021; Chen et al., 2021; Zhang, Wang, & Li, 2021; Marín-Jiménez, Castro, Delgado-Escaño, Kalogeiton, & Guil, 2021; Han, Li, Zhao, & Shen, 2022; Zheng et al., 2022; Song, Huang, Wang, & Wang, 2022; Chai, Li, Zhang, Li, & Wang, 2022; Khan, Farid, & Grzegorzek, 2022; Yu et al., 2022; Liang et al., 2022; Ma, Fu, Zheng, Cao, Hu, & Huang, 2023a; Li & Zhao, 2023). The prevailing body of research predominantly centers on addressing the challenge of cross-view and cross-condition gait recognition, wherein the difficulty lies in establishing feature correlations across disparate viewpoints and walking conditions of the same individual. Within supervised learning frameworks, a substantial portion of existing research begins by enhancing structural complexity to improve feature extraction (Arshad et al., 2020; Khan et al., 2022; Chai et al., 2022; Li & Zhao, 2023; Ma et al., 2023), and researchers have also improved loss functions and training methodologies (Liang et al., 2022; Su et al., 2020; Weichen, Hongyuan, Huang, & Wang, 2022; Zhang, Luo, Ma, Liu, & Li, 2019), aiming to advance the state-of-the-art on public datasets.

In practical application, a gait recognition-based tracking system needs to meet two essential requirements.

1) It should be capable of achieving unsupervised retrieval without relying on manually annotated training datasets. The importance of reducing reliance on manually annotated data is evident for following reasons. Firstly, manual annotation is time-consuming and labor-intensive. If applications of GRT in a new domain requires extensive data labeling, it will increase the cost of using the technology, hindering its widespread adoption and broader application (Ren et al., 2023; Ma et al., 2023). Secondly, using pre-existing annotated datasets to train supervised models can be unsatisfactory when there's a discrepancy between the training data and real-world testing data (Zheng et al., 2021; Ren et al., 2023). For instance, if the training data was collected with cameras positioned low, but in actual application, cameras are mounted at higher positions, there will be a bias between the test and training data. There are researches investigating the unsupervised gait recognition which will be introduced in the section of related work.
2) The system must achieve a top-1 retrieval accuracy rate very close to 100 %; otherwise, the path tracking of some individuals will be erroneous. Additionally, it should clearly distinguish between the distances of top-1 and top-2 retrievals for all samples. By setting a threshold, this ensures the system can accurately identify and respond, regardless of the target's presence in the retrieval database.

However, extensive existing supervised methods (Wan, Wang, & Phoha, 2018; Wu, & Xu, 2019; Zhang et al., 2020; Su et al., 2020; Arshad et al., 2020; Chao et al., 2021; Chen et al., 2021; Zhang et al., 2021; Marín-Jiménez et al., 2021; Han et al., 2022; Zheng et al., 2022; Song et al., 2022; Chai et al., 2022; Khan et al., 2022; Yu et al., 2022; Liang et al., 2022; Ma et al., 2023; Li & Zhao, 2023) along with the few unsupervised approaches available (Zheng et al., 2021; Wang et al., 2022; Ma et al., 2023; Ren et al., 2023) are unable to be implemented in real-world application due to the reliance on the pre-labeled training set and the insufficient retrieval accuracies. This undertaking presents a challenge.

Fortunately, within the realm of electronic tracking systems, challenges such as cross-view matching, which is a major focus in scholarly research, can be effectively mitigated by strategically positioning cameras along narrow pathways. Nature reserves with intricate terrains often feature limited access points, typically through one or a few narrow paths. For instance, despite the vast expanse of Cangshan Mountain, which covers hundreds of square kilometers, entry is restricted to a few hundred narrow paths due to the complex terrain. Placing cameras above these constrained and level surfaces enables the capture of individuals' frontal views at a same-degree angle, say 0-degree, see Fig. 1-b. Consequently, the primary focus of the tracking system revolves around matching individuals within the same viewpoint, eliminating the need for cross-view matching, which necessitates supervised methods. To achieve same-view gait recognition through unsupervised approaches is feasible.

In this study, we introduce a novel and practical electronic tracking system specifically designed for large-scale nature reserves situated in complex terrains. This system is built upon a newly and simple proposed unsupervised gait recognition method. The core technique involves the utilization of a large-scale vision transformer (VIT) model (Han et al., 2023), which is pretrained on the ImageNet dataset (Deng et al., 2009), coupled with a modified contrastive learning strategy, where the contrastive learning approach has transitioned from a semi-supervised (Schiappa, Rawat, & Shah, 2022) to a fully unsupervised method by altering the selection of positive samples.

The pre-trained ViT adeptly extracts relevant features from individual gait data samples. Concurrently, contrastive learning enhances the model's ability to differentiate between individuals entering the reserve, thus providing an ideal representation of each subsample where essential information is encapsulated and redundant information is discarded. Consequently, the proposed unsupervised gait recognition-based tracking system has the capability to track the paths of individuals by correctly retrieval the target from each camera, detecting instances when someone enters but does not exit, exits but does not enter, or enters a restricted area. Furthermore, the systems necessitate only a small amount of binary-processed, low-quality gait data transmission, reducing their reliance on network and resource capacity. This approach adeptly tackles the challenges of large-scale nature reserves with complex terrains, limited power resources, and high bandwidth network infrastructure, all while minimizing maintenance costs. It enables precise control and effectively mitigates the impact of human activities on protected areas at a low cost. The contributions of this research can be summarized as follows:

- The approach introduces a simple yet effective unsupervised gait recognition method that consistently achieves 100 % accuracy in top-1 retrieval for same-view matching, outperforming existing methods. It also ensures a distinct separation between the distances of top-1 and top-2 retrievals for all samples, demonstrating robust identification and response capabilities, even when the target's presence in the retrieval database varies
- In a groundbreaking endeavor, the gait recognition method is applied to the field of environmental conservation, providing an electronic tracking system solution for large-scale nature reserves with complex terrains. This innovative solution can significantly reduce the management costs of natural reserves and, more importantly, mitigates the risk of human interference in these protected areas. Yunnan, China, for instance, houses over 100 nature reserves, highlighting the practical relevance of our work. This research addresses a gap by applying the advanced gait recognition technology

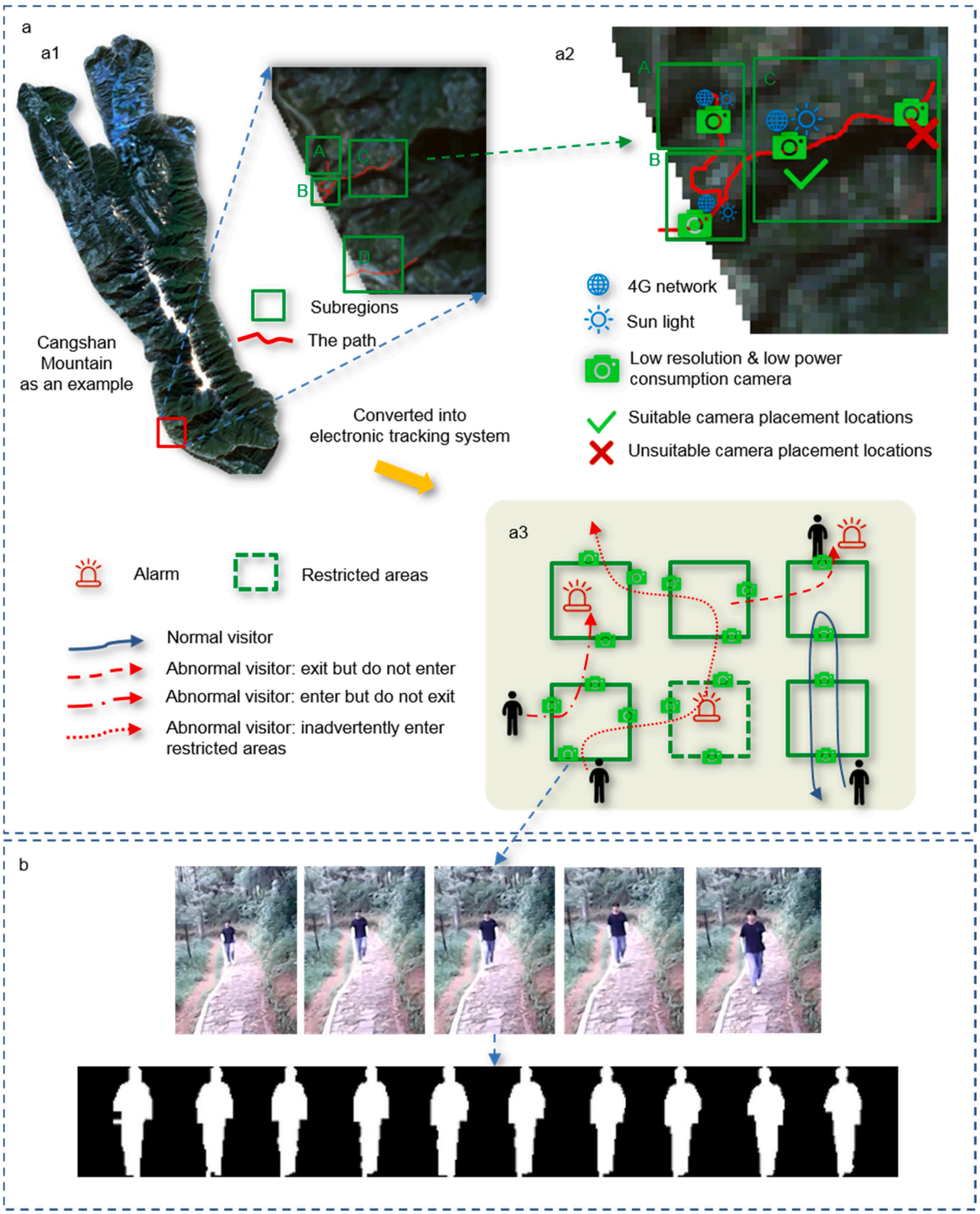


Fig.1. The illustration of tracking system. Panel-a. The key idea is to divide the nature reserves into subregions and focus on the main path in and out this region. The tracking is done by retrieval the individuals from different cameras that record their gait data. Panel-b. Example of the gait images. By positioning the camera directly above the pathway, gait data of pedestrians can be captured from a same-angle view, here, 0-degree angle is taken as an example.

to the urgent task of supervising human activity in expansive nature reserves.

2. Review of related literature and initial considerations

Gait, compared to other biometric features, has many advantages, making GRT a promising area with broad application prospects. This field has always been a research hotspot, with numerous studies exploring the topic (Sepas-Moghaddam & Etemad, 2022; Singh et al., 2018). This chapter will review representative research in the field, showcasing the development GRT and highlighting the core contribution of our work: Unlike existing academic research, our study is the first to apply GRT to the conservation of natural reserves, providing a rare case for the practical application of GRT in real-world scenarios. Based

on the rich achievements in gait research and the exemplary role of our work, we believe that in the future, more research will apply advanced artificial intelligence technology to practical problems.

The quintessential challenge in gait recognition is the precise extraction of individual-specific traits from raw gait data (Sepas-Moghaddam & Etemad, 2022; Singh et al., 2018). It involves discerning the unique aspects of a person's walk, i.e., those idiosyncratic patterns that distinguish one individual from another. These distinguishing features, such as the particular manner in which a person's toes strike the ground or the swing amplitude of their arms, are encapsulated within the gait silhouette or its temporal progression. The challenge for algorithms lies in the nature of the input data: a vast array of numerical values that, while rich with vital information, are also intermingled with superfluous data. Consequently, unlike humans, models do not innately recognize these pivotal features. Despite the extensive body of research on gait recognition, the focal point consistently circles back to feature extraction, that is, developing sophisticated algorithms capable of isolating the essential features from the gait data, filtering out the noise, and ensuring accurate identification of individuals.

2.1. Algorithms designed based on prior knowledge

Algorithms that extracted features manually based on prior knowledge dominated the early stages of gait recognition. The core idea of these methods was to utilize the prior knowledge of the physical motion characteristics of the human body and extract gait features through manually designed algorithms. Representative methods are mainly divided into two categories: model-based and model-free approaches. Model-based approaches rely on a predefined human body model, typically skeletal or joint models, to capture the dynamic characteristics of gait. For instance, the stick figure model (Nixon, Carter, Cunado, Huang, & Stevenage, 1996) uses a simplified skeletal model to represent the human body and identifies gait by analyzing joint movements. In contrast, model-free methods extract features directly from video data without relying on a predefined human body model. For example, the gait energy image (Han & Bhanu, 2005) is a commonly used model-free method that captures the primary features of gait by creating an image through averaging the silhouette of a person walking.

2.2. Supervised methods based on deep neural networks

Early gait feature extraction methods, which relied on manually designed algorithms, did not require parameter updates through supervised or unsupervised training. This approach made them less data-dependent and computationally economical. However, their effectiveness was significantly limited by the algorithms' design. In particular, they often lacked robustness under variable conditions, such as changing environments (Sepas-Moghaddam & Etemad, 2022; Singh et al., 2018).

The development of manually annotated gait datasets has catalyzed a shift in gait feature extraction towards data-driven approaches. Notable datasets include the CASIA series with various scales and conditions (Andrie, Basuki, & Arai, 2011), the large-scale OU-MVLP for experimental settings (Takemura, Makihara, Muramatsu, Echigo, & Yagi, 2018), and outdoor datasets like GREW (Zhu et al., 2021) and Gait 3D (Seely, Samangooei, Lee, Carter, & Nixon, 2008). These datasets have enabled the use of supervised learning methods, particularly convolutional neural networks (CNNs), which leverage Gait Energy Images (GEI) to learn complex features from extensive labeled data (Yan, Zhang, & Coenen, 2015). With known pairings of individuals in the training set, CNNs excel at extracting features that identify the same person across various views and conditions, thus significantly improving cross-view matching in gait recognition (Shiraga, Makihara, Muramatsu, Echigo, & Yagi, 2016).

Compared to manually designed algorithms, deep learning-based methods are more effective and robust in feature extraction. These methods significantly reduce the reliance on prior gait knowledge, thereby also reducing the complexity of implementation. As a result, deep learning has been widely applied in gait recognition research, especially in addressing complex cross-view and cross-condition recognition challenges. These methods have further refined data processing, network structures, and training strategies. For instance, generative adversarial networks (GANs) have been introduced for gait feature extraction (Yu, Chen, Garcia Reyes, & Poh, 2017); Coupled patch alignment achieved cross-view gait matching by locally aligning gait images (Ben et al., 2019); and cross-view gait recognition methods based on classification loss and distance loss have been introduced (Han et al., 2022). Over time, a plethora of classic supervised methods based on deep learning have emerged, such as GaitEdge (Liang et al., 2022), GaitStrip (Wang et al., 2022), and OpenGait (Fan, Liang, Shen, Hou, Huang, & Yu, 2023), progressively pushing the state-of-the-art on public datasets to new heights.

2.3. Application-oriented and unsupervised methods

To address challenges in real-world applications, academic research began to explore new challenges and corresponding solutions. For instance, the UGaitNet method was proposed to tackle missing gait data issues in practical scenarios (Marín-Jiménez et al., 2021). Beyond image-based data, the field has also seen the introduction of unsupervised methods and sensor-based applications (Huitzil, Dranca, Bernad, & Bobillo, 2019).

With deeper consideration of numerous potential application scenarios, it became evident that to truly implement GRT in real-world settings, there's a need to move away from the supervised learning paradigm. Although supervised methods perform well, have low dependency on prior knowledge, and are straightforward in algorithm design, they heavily rely on training datasets with labeled guidance for model learning. Manual annotation is both time-consuming and expensive, hindering the promotion and application of GRT in new fields (Ren et al., 2023; Ma, Fu, Zheng, Peng, Cao, & Huang, 2023b). On the other hand, using existing manually annotated datasets as training sets also poses challenges. For instance, if there's a discrepancy between real-world test data and training set data, the performance of models derived from supervised methods can be affected (Zheng et al., 2021; Ren et al., 2023).

In contrast, unsupervised learning methods, which don't rely on labels, have been proven to achieve feature extraction results and robustness comparable to, or even surpassing, supervised methods (Dai et al., 2023; Fang et al., 2023; Gao et al., 2023; Zhou, Gu, Fang, & Lin, 2022). Thus, unsupervised identity retrieval techniques based on deep learning have gradually become a new research hotspot. For instance, there's a plethora of research on unsupervised re-identification based on single full-body picture (Lin, Wang, & Liu, 2021; Lv, Chen, Li, & Yang, 2018; Xuan & Zhang, 2021). However, unsupervised gait recognition research based on deep learning is still in its infancy (Ren et al., 2023), with existing unsupervised methods primarily focusing on domain adaptation studies (Zheng et al., 2021). Examples include adapting labeled gait identifiers from indoor scenarios to outdoor and wilderness settings (Wang et al., 2022; Ma et al., 2023) or exploring how pretrained gait recognition models can adapt to unlabeled datasets (Ren et al., 2023).

2.4. The gap between practical applications and academic research

GRT has seen significant progress, with extensive research into data processing and both supervised and unsupervised methods. Its potential to enhance forensic, security, immigration, and surveillance systems is vast. Yet, there's a disconnect between academic focus and practical application challenges. This misalignment hinders the real-world deployment of gait recognition, causing a lag in practical use compared to academic advancements (Makihara, Nixon, & Yagi, 2020).

For instance, this study, aimed at natural conservation applications, reveals three key mismatches between academic pursuits and field necessities.

1) Academic efforts are heavily geared towards solving cross-view matching in gait recognition (Ben et al., 2019; Fan et al., 2023; Han et al., 2022; Liang et al., 2022; Wang et al., 2022; Yu et al., 2017). However, in real-world conservation scenarios, cross-view may not be as critical. Practical solutions, such as specialized hardware setups, can effectively bypass the need for complex cross-view algorithms.
2) For practical deployment, the tracking algorithm must consistently deliver a retrieval accuracy rate very close to 100 % for top-1 matches and clearly differentiate between top-1 and top-2 matches. This precision is crucial for correctly identifying targets, whether they are present in the database or not. Yet, even with their advanced accuracy, both supervised and current unsupervised algorithms (Yu et al., 2017; Ben et al., 2019; Han et al., 2022; Liang et al., 2022; Wang et al., 2022; Fan et al., 2023; Ren et al., 2023) fall short of this rigorous standard.
3) In practical applications within nature reserves, establishing a database containing all potential individuals entering the reserve is challenging, making real-time retrieval infeasible. Instead, offline retrieval is a more suitable choice. This means that after collecting a certain amount of data, a unified search can be conducted for a specific individual using data collected from each camera, rather than immediately searching every time new data is collected. However, current academic research has not highlighted the significant advantage of offline retrieval for unsupervised methods compared to supervised methods. For instance, unsupervised methods can fully incorporate test data into encoding training. This strategic use of available data can significantly improve accuracy in applications and is essential. Yet, existing unsupervised methods mainly focus on comparing with supervised methods under the same standards, such as, like supervised methods, only using a portion of the data for encoding training and not incorporating test data into encoding training (Ren et al., 2023). Although for online retrieval tasks, it's challenging to involve the test set in encoding training, for offline retrieval tasks, such a strategy results in generally low accuracy for unsupervised methods, leading to the misconception that unsupervised methods are not suitable for practical applications.

Lastly, it's worth noting that existing academic research primarily focuses on optimizing metrics on public datasets. Many studies overlook the introduction of the methodology and principles, meaning they don't emphasize explaining to readers why the proposed method works effectively. This trend makes neural networks easily perceived as an opaque "black box," with their actual working mechanisms not receiving adequate attention. If researchers could approach from a practical application perspective, elucidate the key challenges, and focus on introducing the core ideas to address these critical issues, providing detailed technical constructions based on these ideas, it would be more enlightening for future explorations that bridge academia and practical applications.

2.5. Initial considerations of our work

This study revolves around the specific issue of personnel monitoring in natural reserves. The key challenge is to consistently achieve a top-1 retrieval accuracy very close to 100 % and to have a clear demarcation between top-1 and top-2 retrievals.

Our approach to address this challenge is as follows:

1) We leverage the unique terrain of natural reserves. By strategically placing cameras, we obtain gait data from the same viewpoint, thereby avoiding the challenge of cross-view matching.
2) We ingeniously take advantage of the benefits of offline retrieval. We make full use of all collected data and adopt a two-step feature extraction technique for gait data. This involves using a pre-trained large model on images to extract features from individual gait data, followed by a differentiated comparison of all gait data to further eliminate redundant information and enhance differentiated information. Under this approach, we simply use the out-of-box method to achieve our goal, see method section.

Our work is a typical example of implementing a specific application based on unsupervised GRT. We believe it can initiate a new research trend of applying the vast array of developed AI technologies to practical applications.

3. The main method

3.1. Problem description: The idea of the track system

The proposed tracking system ingeniously leverages the challenges posed by complex terrains and converts them into advantageous opportunities. As mentioned, nature reserves with intricate terrains often boast restricted access points, one subregion typically confined to one or just a few narrow pathways. Taking advantage of this characteristic, one can divide the whole conservation area into sub-regions, with each subregion having only one or two small paths for entry and exit. Then, one can strategically install our low-resolution and low-power-consumption cameras at critical intersections where 4G signal and solar panels are available (refer to Fig. 1-a-1 and 1-a-2). This strategic positioning allows us to track the movements of individuals as they enter the corresponding sub-regions of the natural reserves, serving multiple purposes. For example, it enables us to detect abnormal individuals who enter but fail to exit, exit but do not enter, or enter restricted areas, see Fig. 1-a-3. Ultimately, this approach effectively mitigates the impact of human activities on the protected area. For instance, if an individual is identified as not having left, an emergency rescue can be launched at night, targeting their last recorded location.

An example of the gait data from one camera installed in the Cangshan Mountain is given in Fig. 1-b. It reflects the typical characteristics of entry paths into a natural reserve with complex terrains. By strategically selecting camera placement locations, one can capture gait video of individuals from a same-degree angle, here 0-degree angle is presented. Therefore, the cross-view matching is not the major concern here. By applying the simple image processing algorithm, such as frame differencing (Jia, Wang, & Li, 2015), one can obtain binary gait image sequences. which occupy only a small amount of space after compression. For instance, this process transforms original 640x480 resolution color images from approximately 130 KB to nearly 5 KB after binarization and cropping, reducing network bandwidth demands. Alternatively, transmitting low-resolution mp4 videos proves efficient, e.g., a 640x480 resolution video at 10 fps and 10 s long amounts to a mere 650 KB.

3.2. The proposed unsupervised gait recognition method

In traditional supervised learning, models are guided to learn essential features within sample data by means of pre-annotated label information. However, the previous work discovered that beyond supervised learning methods, unsupervised learning methods can also steer models to comprehend crucial information within sample data (Zhou et al., 2022). Moreover, unsupervised methods circumvent biases induced by labeled training sets, often outperforming supervised methods across diverse tasks (Dai et al., 2023; Fang et al., 2023; Gao et al., 2023; Zhou et al., 2022).

Our proposed method distinguishes itself from existing unsupervised GRT (Wang et al., 2022; Ma et al., 2023; Ren et al., 2023), primarily in following aspects: it employs an offline matching strategy. Take

Cangshan as an example: the chilly nights deter tourists from staying overnight. As such, our data collection begins at 5 a.m. and wraps up by 6p.m. Once collected, the strategy involves using all the data from the test set for encoding training incorporating an effective two-step feature extraction process. 1) Extracting latent features from individual gait data to filter out noise and redundancy within individual samples; 2) Leveraging differences among various gait data in the contrastive dataset to retain and enhance significant features within these latent features while discarding less important ones. Through these two steps (see Fig. 2-a), efficient encoding of each sample is attained, resulting in a closer proximity of gait data from the same individual in the feature space, ultimately leading to efficient retrieval (see Fig. 2-b). The specific method is implemented by using out-of-box method, as shown in follows:

1-I) The individual's video is converted into gait images through a simple procedure, yielding usually more than 40 binary gait images per person. For a person A, we randomly select 40 images as the individual's data, characterized $\{x_1^A, x_2^A, \cdots, x_{40}^A\}$.

1-II) The pretrained (VIT) without any fine-tune is employed to encode each gait image, resulting in a collection of 768-dimensional vectors, denoted by $\{e_1^A, e_2^A, \cdots, e_{40}^A\}$ with $e_i^A = VIT(x_i^A)$.

1-III) Next, by setting learnable weights w_i^A, weighted summation is performed on the 40 vectors, $\{e_1, e_2, \cdots, e_{40}\}$, to obtain a 768-dimensional vector d^A, that is

$$d^A = \sum_{i=1}^{40} w_i^A e_i^A \tag{1}$$

1-IV) A single-layer multilayer perceptron (MLP) is applied to map the 768-dimensional vector d^A to a 1,000-dimensional vector f^A. The hidden layer size is 1280 and the activation function is Rectified Linear Unit (Relu).

2-I) Within a separate random selection of 40 gait images, denoted as $\{x'^A_1, x'^A_2, \cdots, x'^A_{40}\}$, obtained through the same process for the same individual under the same camera, the resulting feature vectors f_p^A are considered as positive samples, while treating all other samples, denoted as f^B, as negative samples. For example, if 600 samples are collected, corresponding to 100 individuals, each person may appear in a varying number of cameras. For sample A, the other 599 samples are considered as its negative samples.

2-II) A fully unsupervised-learning based contrastive loss is constructed as

$$loss = -\sum_A \log\left[\frac{\exp(dis(f^A, f_p^A)/\tau)}{\sum_{A\neq B}\exp(dis(f^A, f^B)/\tau) + \exp(dis(f^A, f_p^A)/\tau)}\right] \tag{2}$$

where τ is a hyper parameter, $dis(f^A, f^B)$ is the cosine distance between negative sub-samples of A and B, and $dis(f^A, f_p^A)$ is the distance between A and its positive sub-samples. By adjusting the weights w_i^A and the

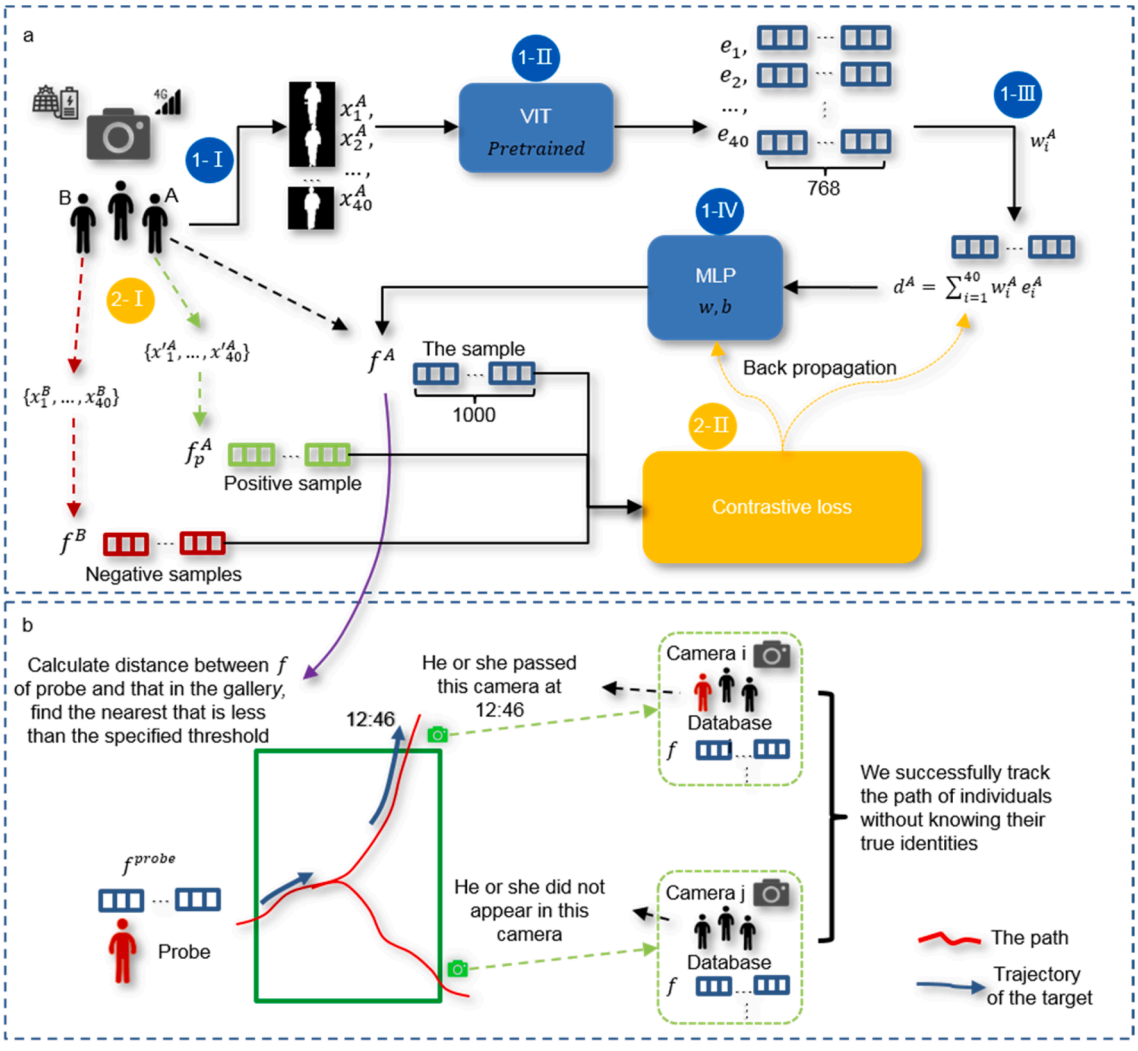


Fig. 2. Panel-a. The illustration of the unsupervised gait recognition method. Panel-b. The retrieval process of the tracking system.

parameters in MLP, the contrastive loss is minimized.

The f^A serves as the ultimate representation for each individual, as shown in Fig. 2-a. It is the key in the downstream retrieval task. For the sake of clarity, f is the named as the final encoding vector. Based on the final encoding vector f of each sub-sample, the system can retrieval the given probe from the galleries, see Fig. 2-b.

4. The results

4.1. The explanation of the experiment and criteria of the evaluation

The dataset. To assess the effectiveness of our method, we conducted experiments using the publicly available CASIA-B dataset (Andrie, Basuki, & Arai, 2011). This dataset consists of gait data collected from 124 individuals under various conditions, such as carrying backpacks. Gait data was captured from 11 different perspectives for each individual, with each perspective recorded independently six times. In typical scenarios, the daily influx of non-tourist personnel entering the protected area through non-tourism routes is approximately a hundred individuals. Therefore, the CASIA-B dataset provides a comprehensive validation of the efficacy of our method and enables meaningful comparisons with other existing approaches.

Exclude cross-view and cross-condition matching. As mentioned in the introduction and method sections that challenges like cross-view matching can be effectively addressed through the strategic placement of cameras along narrow pathways. Therefore, our primary focus revolves around matching individuals within the same viewpoint, thereby excluding cross-view matching. While it is possible for individuals to exhibit varying conditions while traversing the area, our initial data collection indicates that the majority of individuals do not alter their clothing or change their methods of carrying backpacks. As a result, matching across different conditions is not currently considered. However, this aspect will be further explored by integrating gait analysis with other recognition methods based on biosignatures, such as full-body images and obscured facial images.

Off line 1v1 matching. To evaluate the accuracy of matching a specified probe in databases of other cameras, experiments involving 124 individuals walking from a 0-degree angle (other 10 different perspectives were evaluated similarly) are conducted. In this evaluation, six different cameras are utilized to capture a total of 744 gait samples from these individuals. After performing the proposed unsupervised encoding on all the 744 gait samples. Our evaluation process involves selecting individuals from one camera as the probe and matching them against individuals in the other one camera. This matching pattern is referred to as off line 1v1 matching. There are also other matching modes, such as 2v4 matching refers to a scenario where the system uses gait data from the same individual captured by two different cameras as probes, while it searches through the gallery comprising of four other cameras. Compared with other matching modes, 1v1 matching poses the highest level of complexity and is also the required matching mode in real-world applications.

As a result, there are 3,665 individual matches. These matches involve six probe cameras being compared respectively against five gallery cameras, with each match consisting of no more than 124 individual comparisons (note that some cameras had fewer individuals than the total of 124 due to insufficient gait images).

Interpretation of correct retrieval. The algorithm's performance is determined based on whether the correct match for the probe (e.g., individual A) is identified as the first candidate by the algorithm in the other cameras, i.e., the top-1 candidate. In other words, a match is considered accurate if the algorithm correctly identified the probe by the top-1 candidate, while an incorrect identification was considered if it failed to match the probe. Additionally, in cases where the gallery did not include the target person, the algorithm is considered correct if it provided no candidates. This is achieved by setting a threshold: only candidate with distances between themselves and the probe lower than this threshold will be recommended.

The code and computer setups. The code is given in Code Ocean (https://codeocean.com/capsule/9670283/tree/v1), where details of the experiment setups can be found. The experiment is run on computer with CPU i9-13900 k and single NVIDIA GPU 4080.

4.2. The main result

The overall accuracy is evaluated by calculating the number of correct retrievals among the 3,665 individual matches. In order to show the effectiveness of the proposed method, the top-1 and top-2 distances of each retrieval are plotted in Fig. 3-a. It shows that, there is a clear boundary between the top-1 and top-2 distances. Therefore, the experimental results revealed that the top-1 matching accuracy is 100.00 %. The effectiveness of the encoding procedure is illustrated in Fig. 3-b, where the t-SNE visualization of 30 randomly selected individuals each with 6 camera views before and after encoding is given. It shows that, the proposed method successfully identifies the key features of each individual pulling the same individuals together in the feature space.

It demonstrates the effectiveness of our algorithm and supports its application in implementing an electronic tracking system. Notably, it is important to highlight that the method's performance is not contingent on the number of cameras used. Each camera's matching process is independent, enabling the method to achieve close to 100 % matching accuracy even with a larger number of cameras. This observation is illustrated and analyzed in Fig. 3.

4.3. The comparison with existing approaches and further analysis

Comparison with existing approaches. To demonstrate the superiority of our method, a comparison with existing approaches is conducted. Most studies focus on cross-view and cross-condition matching, but our work is dedicated to same-view matching. To our knowledge, only countable supervised methods detail same-view matching results for specific angles (Liao et al., 2020, 2021; Zhang et al., 2019), others report average accuracies, like 98.9 % in normal mode for supervised (Marín-Jiménez et al., 2021) and 90.3 % for unsupervised (Ren et al., 2023) method. We benchmark our results against those studies that provide detailed same-view matching results for particular angles.

The existing supervised learning necessitates prior knowledge of the correspondence between 62 out of the 124 individuals across different cameras to train the models. The testing phase involves a dataset of 62 individuals, with two cameras serving as probes and data from four cameras used as the gallery, i.e., 2v4 matching mode (Zhang et al., 2019;). Here, we adopt the same testing standards using the 2v4 mode with a test set of 62 subjects for our method. This is easier than the 1v1 matching proposed in the main result section.

The results, depicted in Fig. 4, illustrate that our method achieves a 100 % top-1 matching accuracy under the same conditions at the view of all 11 angles, surpassing the performance of existing supervised methods (Liao et al., 2020, 2021; Zhang et al., 2019).

The ablation experiments. Furthermore, our experiments demonstrated the critical roles played by both large-scale model encoding and contrastive learning in our proposed method, as shown in Fig. 5-a. The large-scale model encoding improves the matching accuracy by approximately 71.20 % (from 22.80 % to 92.00 %), while contrastive learning contributes an additional 8.00 % improvement to the final matching accuracy.

Further analysis. The relationship between accuracy and the numbers of gait images selected for each sample is investigated, as shown in Fig. 6-a. It shows that the method needs enough number of gait images to identify individuals correctly. Although, the method achieves top-1 retrieval accuracy of 99.81 % by using 20 gait images, it needs more time to train. To strike a balance between effectiveness and efficiency, 40 gait images is selected. For instance, processing 40 gait images for each sub-sample, our method requires around 4,000 s, just over

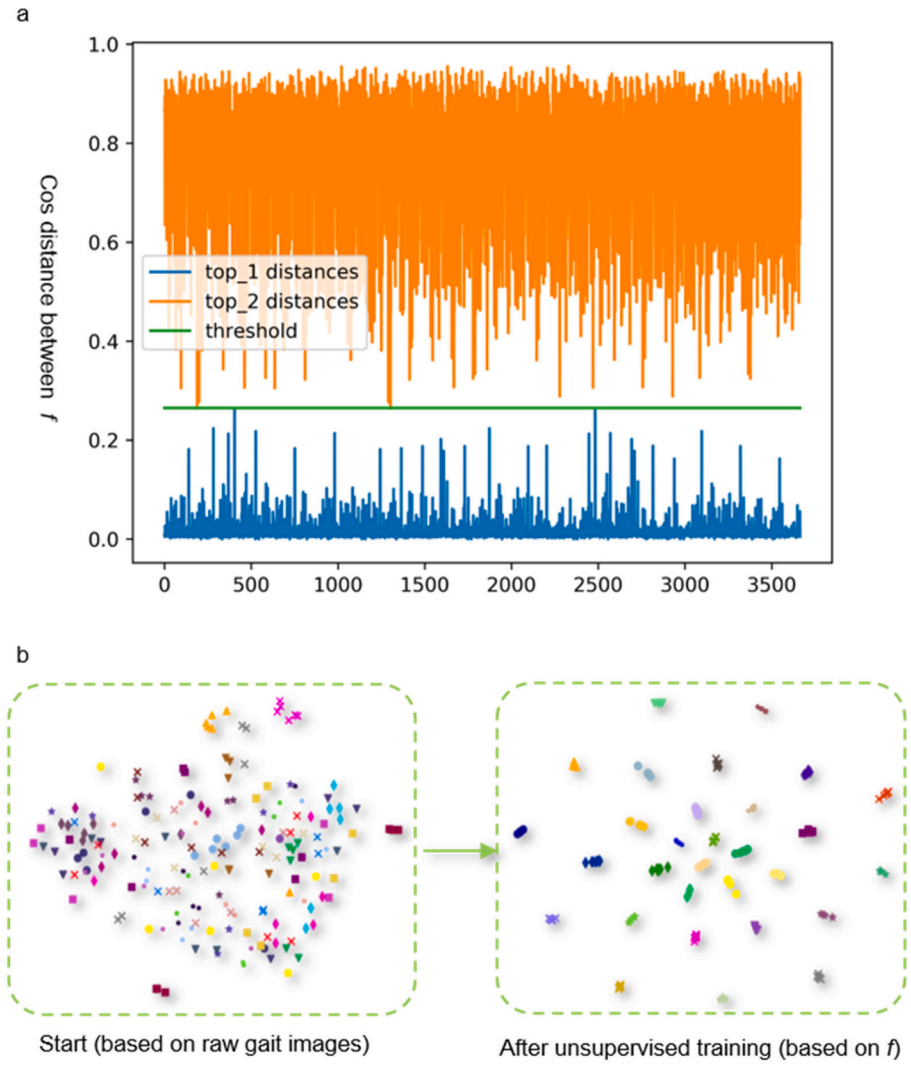


Fig. 3. The illustration and analysis of the result, taking 0-angle view as an example. Panel-a. The top-1 and top-2 distances of each matching. It shows that there is a clear threshold between the top-1 and top-2 distances, making the top-1 retrieval accuracy 100%. Panel-b. The t-SNE visualization of 30 randomly selected individuals each with 6 camera views before and after encoding.

an hour, to report the movement paths of over 100 individuals captured by 6 cameras within the reserve for the day. Further discussion of the algorithm's running time is given in the conclusion section.

The relationship between accuracy and the dimensions of the final encoding vector f, the output of MLP, is investigated, as depicted in Fig. 6-b. It indicates that our method consistently achieves consistently 100.00 % matching accuracy under different dimensions of the encoding vectors f. However, the 1,000-dimensional f contains information about the viewing angle. Fig. 6-b-2 illustrates the accuracy of MLP trained using supervised methods based on f as input when predicting gait angles (it takes the final encoded vector f as input and outputs 11 gait angles). The results show that the accuracy of the 1,000-dimensional f is 100 % when predicting the viewing angle, while the accuracy of other dimensions of f rapidly decreases as the dimensionality decreases. In this study, the 1,000-dimensional f is chosen as the final encoding vector, because it comprehensively captures the gait viewing angle information of the samples, enabling us to address certain minor angle cross-view recognition tasks in subsequent applications.

The robustness of the proposed method across dataset with different numbers of individuals is evaluated, see Fig. 6-c. It indicates that our method consistently achieves consistently 100.00 % matching accuracy with a dataset involving up to one hundred individuals.

The robustness of the proposed method under different matching patterns, e.g., matching modes of 1 V1, 1 V2, 1 V3, and 2 V4 are evaluated, see Fig. 6-d. It indicates that our method consistently achieves consistently 100.00 % matching accuracy under different matching patterns.

4.4. Extra experiments on the OU-MVLP dataset

In this section, we delve deeper into the performance of our proposed method by conducting additional experiments on the OU-MVLP dataset (Takemura et al., 2018).

Similar to the CASIA-B dataset, the OU-MVLP dataset is also among the most frequently used datasets for gait verification. Compared to the CASIA-B dataset, it encompasses gait data of over 10,000 subjects, each captured from 14 different angles, with two independent collections for each angle. Given the larger number of individuals in this dataset and its distinct data collection method compared to CASIA-B, it further validates the robustness of our approach in terms of dataset size and varied gait data collection scenarios. Given that our method necessitates more than 40 gait sequence images for each individual under every camera, we filtered out samples from the complete OU-MVLP dataset that didn't meet this criterion. From these filtered samples, we randomly selected datasets containing 100, 200, 300, and 500 individuals. We didn't test on dataset with more individuals because, in practical applications, the

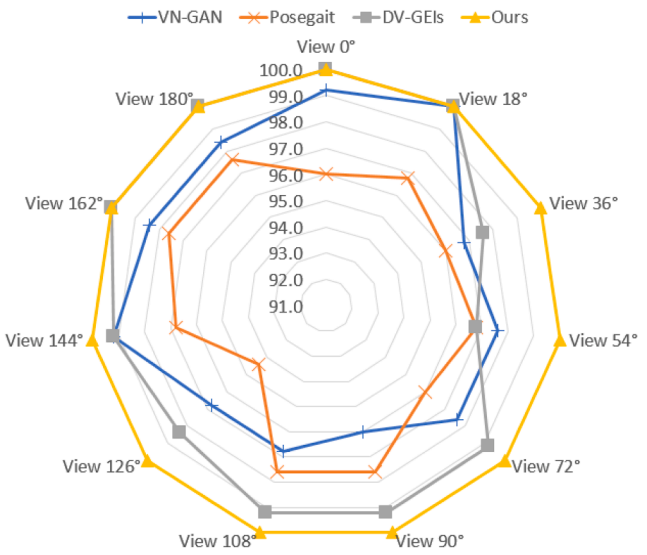


Fig. 4. The result of the same-view matching of our proposed method and that of the existing supervised VN-GAN from (Zhang et al., 2019), Posegait from (Liao, Yu, An, & Huang, 2020), and DV-GEIs from (Liao, An, Li, & Bhattacharyya, 2021). The gait data under normal walking is considered.

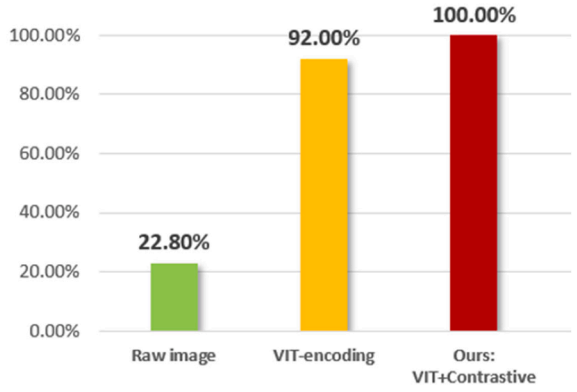


Fig. 5. The results of ablation experiments. It takes 0-angle view as an example.

daily number of people entering each sub-area of the nature reserve through these regulated small paths is around 100. We used the 0-degree angle as a representative.

The final results indicate that our method consistently achieved a 100 % matching accuracy rate in various tests. A notable observation was the clear distinction between the top-1 and top-2 distances across all matching scenarios. This indicates that the proposed electronic tracking system can pinpoint targets in the cameras accurately, and also in scenarios where the target object isn't visible in the camera. This capability ensures precise path tracking, as depicted in Fig. 7.

Furthermore, it's worth noting that, unlike the CASIA-B dataset, the OU-MVLP dataset has instances of incomplete gait data. This incomplete data has a pronounced impact on the accuracy of our method's matching. For instance, if the gait data of individual A, captured under a specific camera, is incomplete, it can adversely affect the retrieval accuracy for A under that camera. As a result, the system might fail to correctly match A's gait, leading to an inability to log the exact path of A. Given this, it underscores the importance of carefully selecting camera installation locations. Further discussion on the improvement of robustness of the system is given in the discussion section.

5. Conclusions and discussions

5.1. Conclusions

Contrary to the extensive GRT research primarily aimed at enhancing cross-view matching accuracy through complex supervised methods, our proposed approach offers a straightforward, unsupervised method tailored for real-world applications where cross-view matching is not a primary concern. It has following advantages, making it highly suitable as the underlying core algorithm for an electronic tracking system in large-scale nature reserves with complex terrains. 1) It needs not any known matching relationships in the training set which is crucial in the application of the tracking system as new individuals enter the area on a daily basis. Moreover, it enables us to track individuals entering the area without relying on a large pre-existing personnel database, thereby reducing expenses associated with maintaining such a database (the only trade-off being the disregard for true individual identities). 2) It achieves consistent 100 % top-1 retrieval accuracies on datasets involving up to over hundreds of individuals, which is the daily influx of non-tourist personnel entering the protected area. Additionally, for all samples, it gives a clear distinction between the distances of top-1 and top-2 retrievals. That is, the method robustly identifies and reacts, regardless of whether the target is present in the retrieval database. As a result, the proposed method leverages the challenges posed by complex terrains and converts them into advantageous opportunities by strategically install cameras at critical intersections. It utilizes only a small amount of binary processed low-quality gait data of each individual and seamlessly integrates with low-resolution and low-power-consumption cameras. These cameras can remain dormant for extended periods and only capture 2–3 s of data when individuals pass by. Instead of transmitting high-quality colorful images, we transmit extremely low-resolution binary gait images, significantly reducing power consumption for data collection and transmission, as well as the demand for network bandwidth.

5.2. Discussions on the limitation

The current tracking system has following limitations. Firstly, due to the use of low-resolution devices, it cannot provide further detailed information about individuals, such as their names. Secondly, it operates as an offline retrieval strategy, generating reports on individuals entering the nature reserve each evening rather than providing real-time tracking. Additionally, the method requires hours to complete the data encoding and retrieval. For example, to give the required final report of the paths of over 100 individuals under 6 cameras into the reserves takes about 4,000 s.

Despite these limitations, we still consider the proposed approach to strike the best balance in terms of cost-effectiveness currently. For example, the proposed method can promptly track and report the paths of non-regular visitors entering Cangshan within a day. Even though there are over 100 entry paths to Cangshan, potentially requiring more than 100 cameras, we've observed that due to the vast terrain, these areas covered by cameras are isolated from each other. These 100 cameras can be further categorized into over 10 distinct zones, with visitors typically only moving within these specific zones. Hence, for 100 visitors, a search time of 1 to 2 h using approximately 10 cameras is deemed acceptable.

We are continuously refining our approach to address the aforementioned limitations. From an algorithmic perspective, we are making the algorithm applicable to the path tracking of more individuals and exploring the creation of large datasets and models specific to gait recognition, and integrating gait analysis with other recognition methods based on biosignatures, such as full-body images and obscured facial images, aiming to achieve a more versatile and stable unsupervised gait recognition technique. On the hardware front, we're investigating the relationship between the algorithm's runtime and hardware

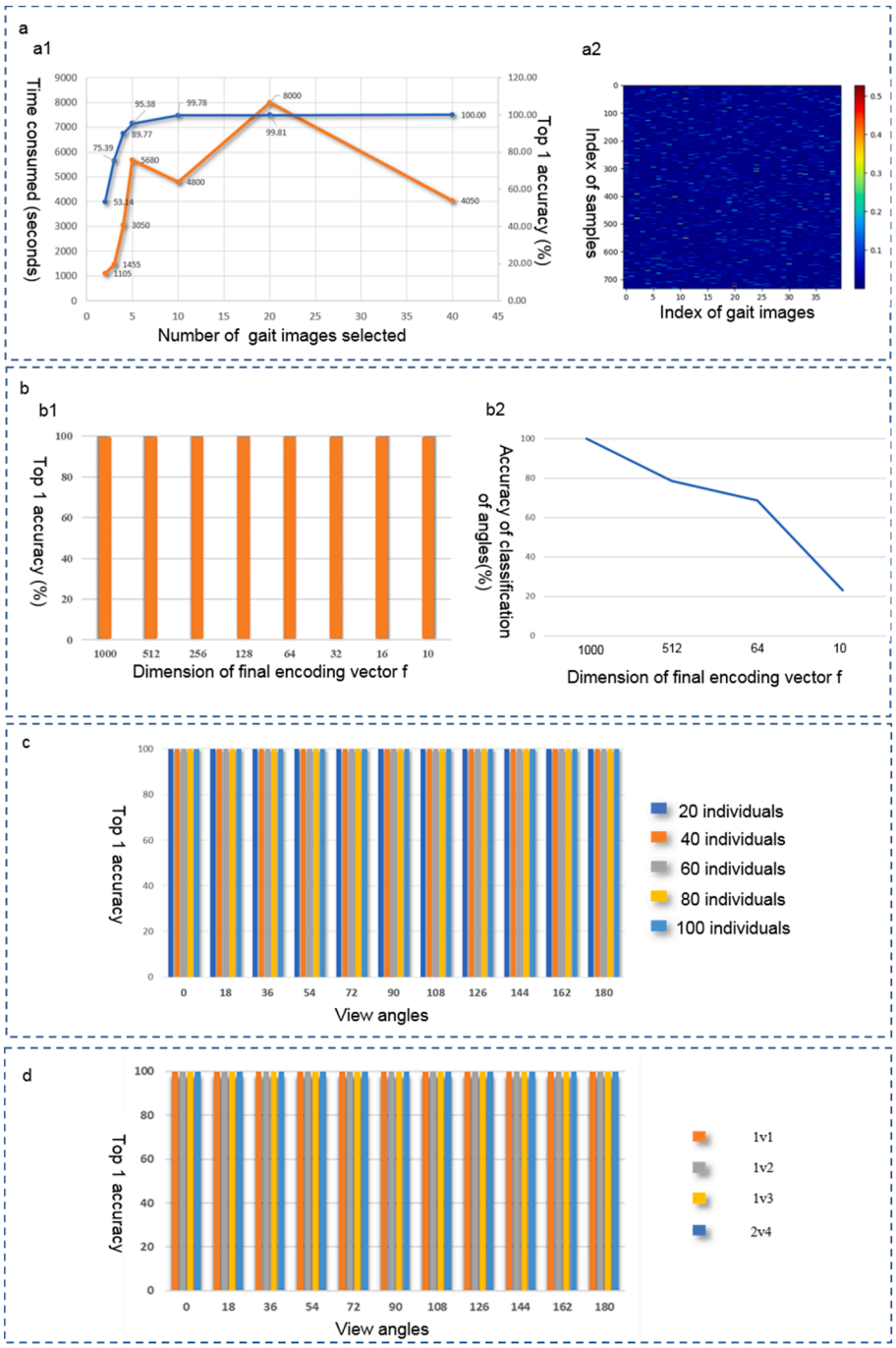


Fig. 6. The results of further analysis. Panel-a. The effectiveness and efficiency of the proposed method over different selections of numbers of gait images. Panel-b. The discussion on the selection of the dimension of the final encoding vector f. Panels-c and -d. The robustness of the proposed method over dataset with different number of individuals and different retrieval patterns.

configurations, seeking an efficient hardware setup that can cater to the conservation needs of other nature reserves.

5.3. Discussions on the significance

Through comprehensive verification, we have demonstrated the effectiveness and stability of this new retrieval technique, establishing its feasibility for cost-effective and efficient personnel supervision in large-scale nature reserves with complex terrains. Looking ahead, we are actively advancing the development of low-cost hardware devices that integrate this new technology. The potential for widespread adoption is significant. Our system offers substantial benefits with minimal construction costs (approximately 1 million RMB, including over 100 low-resolution cameras and associated equipment) and low maintenance costs. By significantly reducing the management and protection expenses of each nature reserve, it effectively addresses the considerable demand, e.g., within and beyond Yunnan Province (which boasts over a hundred natural reserves). The system not only provides significant economic advantages but also contributes to the social well-being of the region. We remain dedicated to its implementation and continuous improvement.

This work is an exemplary case of combining artificial intelligence

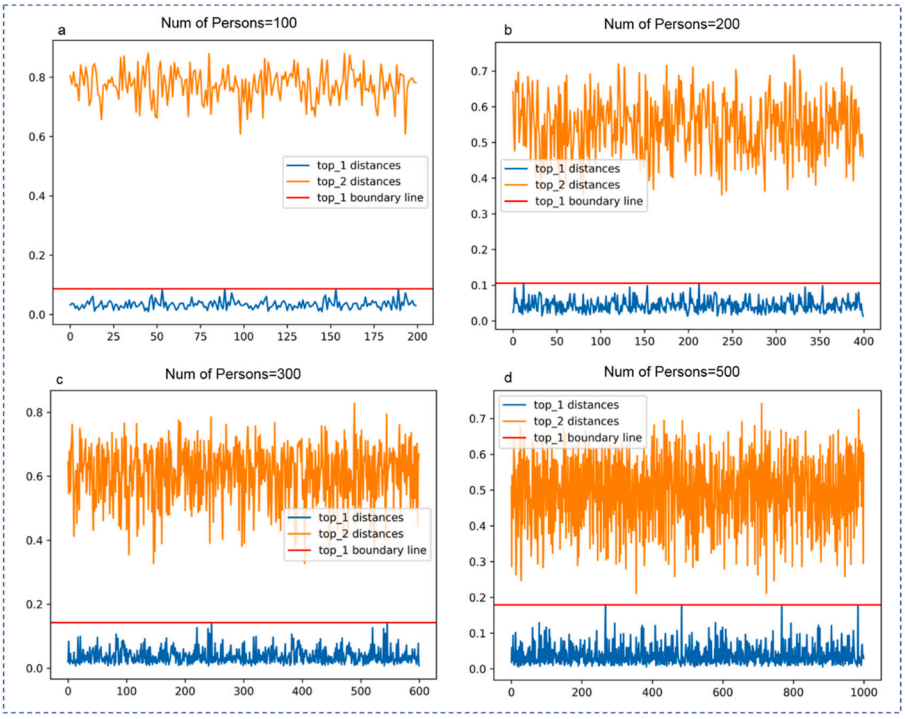


Fig. 7. The top-1 and top-2 distances of each matching for dataset consists of various number of individuals. Panel a. 100 individuals. Panel b. 200 individuals. Panel c. 300 individuals. Panel d. 500 individuals. It shows that there is a clear threshold between the top-1 and top-2 distances, making the top-1 retrieval accuracy 100%.

technology with practical applications in the field of environmental conservation (Chan et al., 2016). It is dedicated to promoting the integration of academic research in artificial intelligence with other areas, such as environmental conservation.

Author contributions statement

Carrying out the experimental work, Xiaolin Guan; Formal analysis, Yao Shen, Zhenyu Zhang, and Junjie Gu; Writing original draft preparation, Xiaolin Guan and Chichun Zhou; Deep learning model development, Xiaolin Guan, Chichun Zhou, and Zhuohang Yu; Analyzing the results, Xiaolin Guan, Zhuohang Yu, Yao Shen, Chichun Zhou, and Zhenyu Zhang; Reviewing the manuscript, Chichun Zhou, Yao Shen, Zhenyu Zhang, and Junjie Gu. All authors have read and agreed to the published version of the manuscript.

CRediT authorship contribution statement

Chichun Zhou: . **Xiaolin Guan:** Formal analysis. **Zhuohang Yu:** . **Yao Shen:** Writing – original draft. **Zhenyu Zhang:** Writing – original draft. **Junjie Gu:** Writing – original draft.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This work was supported by National Nature Science Foundation of China (62106033 and 42367066), Yunnan Fundamental Research Project (202001AU070020), and Yunnan Province Dali Prefecture Science and Technology Bureau, Social Development Field Project (20232904E030002). Yao Shen acknowledges support from Fundamental Research Funds for the Central Universities, China No.2022JKF02024.

References

Andrie, R., Basuki, A., & Arai, K. (2011). A review of Chinese Academy of Sciences (CASIA) gait database as a human gait recognition dataset.

Arshad, H., Khan, M. A., Sharif, M. I., Yasmin, M., Tavares, J. M. R. S., Zhang, Y.-D., & Satapathy, S. C. (2020). A multilevel paradigm for deep convolutional neural network features selection with an application to human gait recognition. *Expert Systems with Applications, 12541*.

Ben, X., Gong, C., Zhang, P., Jia, X., Wu, Q., & Meng, W. (2019). Coupled patch alignment for matching cross-view gaits. *IEEE Transactions on Image Processing, 28*(6), 3142–3157.

Chai, T., Li, A., Zhang, S., Li, Z., & Wang, Y. (2022). Lagrange Motion Analysis and View Embeddings for Improved Gait Recognition. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (CVPR), 20249-20258.

Chan, K. M., Balvanera, P., Benessaiah, K., Chapman, M., Díaz, S., Gómez-Baggethun, E., … Turner, N. (2016). Why protect nature? Rethinking values and the environment. *Proceedings of the national academy of sciences, 113*(6), 1462–1465.

Chao, H., Wang, K., He, Y., Zhang, J., & Feng, J. (2021). GaitSet: Cross-view gait recognition through utilizing gait as a deep set. *IEEE transactions on pattern analysis and machine intelligence, 44*(7), 3467–3478.

Chen, X., Luo, X., Weng, J., Luo, W., Li, H., & Tian, Q. (2021). Multi-view gait image generation for cross-view gait recognition. *IEEE Transactions on Image Processing, 30*, 3041–3055.

Dai, Y., Xu, J., Song, J., Fang, G., Zhou, C., Ba, S., … Kong, X. (2023). The Classification of Galaxy Morphology in the H Band of the COSMOS-DASH Field: A Combination-based Machine-learning Clustering Model. *The Astrophysical Journal Supplement Series, 268*(1), 34.

Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *In 2009 IEEE conference on computer vision and pattern recognition* (pp. 248–255). IEEE.

Fan, C., Liang, J., Shen, C., Hou, S., Huang, Y., & Yu, S. (2023). OpenGait: Revisiting Gait Recognition Towards Better Practicality. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 9707-9716).

Fang, G., Ba, S., Gu, Y., Lin, Z., Hou, Y., Qin, C., … Kong, X. (2023). Automatic classification of galaxy morphology: A rotationally-invariant supervised machine-learning method based on the unsupervised machine-learning data set. *The Astronomical Journal, 165*(2), 35.

Gao, L., Yu, Z., Wang, S., Hou, Y., Zhang, S., Zhou, C., & Wu, X. (2023). A new paradigm in lignocellulolytic enzyme cocktail optimization: Free from expert-level prior knowledge and experimental datasets. *Bioresource Technology, 129758*.

Han, F., Li, X., Zhao, J., & Shen, F. (2022). A unified perspective of classification-based loss and distance-based loss for cross-view gait recognition. *Pattern Recognition, 125*, Article 108519.

Han, J., & Bhanu, B. (2005). Individual recognition using gait energy image. *IEEE transactions on pattern analysis and machine intelligence, 28*(2), 316–322.

Han, K., Wang, Y., Chen, H., Chen, X., Guo, J., Liu, Z., Tang, Y., Xiao, A., Xu, C., Xu, Y., Yang, Z., & Zhang, Y. (2023). A Survey on Vision Transformer. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 45*(1), 87–110.

Huitzil, I., Dranca, L., Bernad, J., & Bobillo, F. (2019). Gait recognition using fuzzy ontologies and Kinect sensor data. *International Journal of Approximate Reasoning, 113*, 354–371.

Isabelle, D. A., & Westerlund, M. (2022). A review and categorization of artificial intelligence-based opportunities in wildlife, ocean and land conservation. *Sustainability, 14*(4), 1979.

Jia, S., Wang, L., & Li, X. (2015). View-invariant gait authentication based on silhouette contours analysis and view estimation. *IEEE/CAA Journal of Automatica Sinica, 2*(2), 226–232.

Khan, M. H., Farid, M. S., & Grzegorzek, M. (2022). *A Comprehensive Study on Codebook-Based Feature Fusion for Gait Recognition*. SSRN.

Li, N., & Zhao, X. (2023). A Strong and Robust Skeleton-Based Gait Recognition Method with Gait Periodicity Priors. *IEEE Transactions on Multimedia, 25*, 3046–3058. https://doi.org/10.1109/TMM.2022.3154609

Liao, R., An, W., Li, Z., & Bhattacharyya, S. S. (2021). A novel view synthesis approach based on view space covering for gait recognition. *Neurocomputing, 453*, 13–25.

Liao, R., Yu, S., An, W., & Huang, Y. (2020). A model-based gait recognition method with body pose and human prior knowledge. *Pattern Recognition, 98*, Article 107069.

Liang, J., Fan, C., Hou, S., Shen, C., Huang, Y., & Yu, S. (2022). GaitEdge: Beyond Plain End-to-End Gait Recognition for Better Practicality. In S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, & T. Hassner (Eds.), *Computer Vision – ECCV, ECCV 2022 Lecture Notes in Computer Science*. Cham: Springer.

Lin, Z., Wang, X., & Liu, J. (2021). Cross-domain person re-identification with adversarial alignment and instance mining. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 43*(1), 122–136.

Liu, W., Wang, B., Yang, Y., Mou, N., Zheng, Y., Zhang, L., & Yang, T. (2022). Cluster analysis of microscopic spatio-temporal patterns of tourists' movement behaviors in mountainous scenic areas using open GPS-trajectory data. *Tourism Management, 93*, Article 104614.

Lv, J., Chen, W., Li, Q., & Yang, C. (2018). Unsupervised cross-dataset person re-identification by transfer learning of spatial-temporal patterns. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 7948-7956).

Ma, K., Fu, Y., Zheng, D., Cao, C., Hu, X., & Huang, Y. (2023a). Dynamic Aggregated Network for Gait Recognition. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (CVPR), 22076-22085.

Ma, K., Fu, Y., Zheng, D., Peng, Y., Cao, C., & Huang, Y. (2023b). Fine-grained Unsupervised Domain Adaptation for Gait Recognition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 11313-11322).

Marín-Jiménez, M. J., Castro, F. M., Delgado-Escaño, R., Kalogeiton, V., & Guil, N. (2021). UGaitNet: Multimodal gait recognition with missing input modalities. *IEEE Transactions on Information Forensics and Security, 16*, 5452–5462.

Nixon, M. S., Carter, J. N., Cunado, D., Huang, P. S., & Stevenage, S. V. (1996). Automatic gait recognition. *Biometrics: Personal Identification in Networked Society*, 231–249.

Pickering, C., Rossi, S. D., Hernando, A., & Barros, A. (2018). Current knowledge and future research directions for the monitoring and management of visitors in recreational and protected areas. *Journal of Outdoor Recreation and Tourism, 21*, 10–18.

Schiappa, M. C., Rawat, Y. S., & Shah, M. (2022). Self-Supervised Learning for Videos: A Survey. *ACM Computing Surveys, 55*(6), 1–39.

Seely, R. D., Samangooei, S., Lee, M., Carter, J. N., & Nixon, M. S. (2008). In *September). The university of southampton multi-biometric tunnel and introducing a novel 3d gait dataset* (pp. 1–6). IEEE.

Sepas-Moghaddam, A., & Etemad, A. (2022). Deep gait recognition: A survey. *IEEE transactions on pattern analysis and machine intelligence, 45*(1), 264–284.

Shiraga, K., Makihara, Y., Muramatsu, D., Echigo, T., & Yagi, Y. (2016). In *June). Geinet: View-invariant gait recognition using a convolutional neural network* (pp. 1–8). IEEE.

Singh, A., Patil, D., & Omkar, S. N. (2018). Eye in the sky: Real-time drone surveillance system (dss) for violent individuals identification using scatternet hybrid deep learning network. In *In Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 1629–1637).

Singh, J. P., Jain, S., Arora, S., & Singh, U. P. (2018). Vision-based gait recognition: A survey. *IEEE Access, 6*, 70497–70527.

Song, C., Huang, Y., Wang, W., & Wang, L. (2022). CASIA-E: A large comprehensive dataset for gait recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 45*(3), 2801–2815.

Su, J., Zhao, Y., & Li, X. (2020). In *Deep Metric Learning Based On Center-Ranked Loss for Gait Recognition* (pp. 4077–4081). IEEE.

Takemura, N., Makihara, Y., Muramatsu, D., Echigo, T., & Yagi, Y. (2018). Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition. *IPSJ transactions on Computer Vision and Applications, 10*, 1–14.

Wan, C., Wang, L., & Phoha, V. V. (Eds.). (2018). A survey on gait recognition. *ACM Computing Surveys (CSUR)*, 51(5), 1-35.

Wang, M., Lin, B., Guo, X., Li, L., Zhu, Z., Sun, J., … Yu, X. (2022). GaitStrip: Gait Recognition via Effective Strip-based Feature Representations and Multi-Level Framework. In *In Proceedings of the Asian Conference on Computer Vision* (pp. 536–551).

Weichen, Y.u., Hongyuan, Y.u., Huang, Y., & Wang, L. (2022). In *Generalized Inter-class Loss for Gait Recognition* (pp. 141–150). New York, NY, USA: Association for Computing Machinery.

Xuan, S., & Zhang, S. (2021). Intra-inter camera similarity for unsupervised person re-identification. In *In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 11926–11935).

Yan, C., Zhang, B., & Coenen, F. (2015). In *October). Multi-attributes gait identification by convolutional neural networks* (pp. 642–647). IEEE.

Yu, S., Chen, H., Garcia Reyes, E. B., & Poh, N. (2017). Gaitgan: Invariant gait feature extraction using generative adversarial networks. In *In Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 30–37).

Zhang, K., Luo, W., Ma, L., Liu, W., & Li, H. (2019). In *Learning Joint Gait Representation via Quintuplet Loss Minimization* (pp. 4700–4709). IEEE.

Zhang, S., Wang, Y., & Li, A. (2021). Cross-view gait recognition with deep universal linear embeddings. In *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 9095–9104).

Zheng, J., Liu, X., Liu, W., He, L., Yan, C., & Mei, T. (2022). Gait recognition in the wild with dense 3d representations and a benchmark. In *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 20228–20237).

Zheng, J., Liu, X., Yan, C., Zhang, J., Liu, W., Zhang, X., & Mei, T. (2021). In *May). Trand: Transferable neighborhood discovery for unsupervised cross-domain gait recognition* (pp. 1–5). IEEE.

Zhou, C., Gu, Y., Fang, G., & Lin, Z. (2022). Automatic morphological classification of galaxies: Convolutional autoencoder and bagging-based multiclustering model. *The Astronomical Journal, 163*(2), 86.

Zhu, Z., Guo, X., Yang, T., Huang, J., Deng, J., Huang, G., … Zhou, J. (2021). Gait recognition in the wild: A benchmark. In *In Proceedings of the IEEE/CVF international conference on computer vision* (pp. 14789–14799).