Human Identification at a Distance: Challenges, Methods and Results on HID 2023

Shiqi Yu[1], Chenye Wang[2], Yuwei Zhao[3], Li Wang[4], Ming Wang[5],
Qing Li[6], Wenlong Li[7], Runsheng Wang[8],Yongzhen Huang[9,10],
Liang Wang[11], Yasushi Makihara[12], and Md Atiqur Rahman Ahad[13]
[1]Southern University of Science and Technology, China. [2]Shandong University, China. [3]Beihang University, China.
[4]Dalian Everyspry Sci&Tech, China. [5]Beijing Jiaotong University, China. [6]Harbin Institute of Technology, China.
[7]Xidian University, China. [8]Huazhong University of Science and Technology, China. [9]Beijing Normal University, China.
[10]Watrix Technology Limited Co. Ltd. [11]Institute of Automation, Chinese Academy of Sciences, China.
[12]Osaka University, Japan. [13]University of East London, UK.
https://hid2023.iapr-tc4.org/

Abstract

Human Identification at a Distance (HID) is an important research area due to its importance (especially in biometrics) and inherent challenges within this domain. To mitigate some of the constraints, we have introduced the HID challenge. This paper presents an overview of the 4th International Competition on Human Identification at a Distance (HID 2023), which serves as a benchmark for evaluating various methods in the field of human identification at a distance. We have introduced a new dataset, SUSTech-Competition, engulfing a cross-domain challenge. This dataset has 859 subjects, having various variations of clothing, carrying conditions, occlusions, and view angles. With a substantial participation of 254 registered teams, HID 2023 has attracted considerable attention and yielded highly encouraging results. Notably, the top-performing teams achieved significantly good accuracies. In this paper, we provide an introduction to the competition, encompassing the dataset, experimental settings, and competition organization, as well as an analysis of the results obtained by the top teams. Additionally, we delve into the methodologies employed by these leading teams. The progress demonstrated in this competition offers an optimistic outlook on the advancements in gait recognition, highlighting its potential for robust real applications.

1. Introduction

Human identification at a distance poses significant challenges, as most traditional biometric modalities are difficult to acquire under such circumstances. However, with the increasing need for enhanced security measures, there is a growing demand for reliable methods of human identification at a distance. Gait recognition has emerged as a promising biometric modality for this purpose, as it can be captured even when faces are obscured or too small to be detected, making it valuable in scenarios where other biometric features are unavailable. Due to the pandemic, where facial masks are commonly worn in public, gait recognition becomes even more relevant as it may be the only viable biometric feature at a distance.

Gait recognition has witnessed significant advancements, particularly with the advent of deep learning techniques, since its inception in the late 1990s. Innovative algorithms, such as GaitSet [2], GaitPart [5], and GaitGL [10], have been developed, showing promising results. However, like many research areas in computer vision, the performance of gait recognition can be influenced by various factors, and different experimental settings can yield different outcomes. Real-world applications of gait recognition often lead to a noticeable decrease in accuracy, as highlighted in recent studies [20, 18]. For instance, according to the results presented in [18], the rank-1 accuracy on the challenging Gait3D dataset is only 53.20%, whereas it can easily surpass 95% on indoor datasets like CASIA-B [17]. These findings demonstrate that gait recognition still has a long way to go before it can achieve the desired accuracy and robustness in real-world scenarios.

To improve gait recognition research and enable fair comparisons and evaluations in complex environments, the International Competition on Human Identification at a Distance has been organized since 2020, with HID 2023 being the fourth edition. The previous competitions have achieved remarkable success, with consistent improvements in recog-

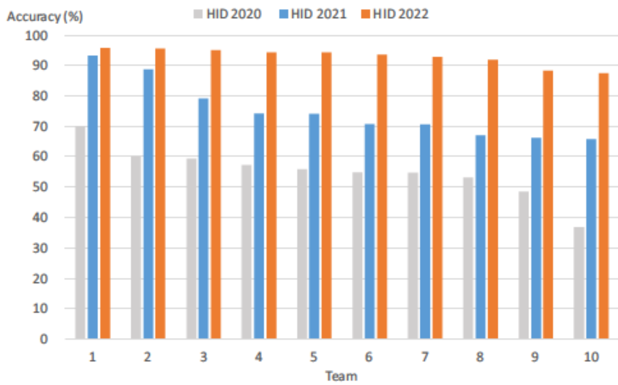

Figure 1. The top 10 results of HID 2020, HID 2021 and HID 2022 [16]. The dataset for the 3 competitions is the same, CASIA-E. The results of HID 2020 and HID 2021 have been calibrated according to the same standard as HID 2022.

nition accuracy each year, as depicted in Figure 1. In HID 2022, the best accuracy reached 95.9%, even though the test set was challenging. It appears that the accuracy on the dataset has almost reached saturation. Therefore, in HID 2023, we introduced a new dataset, SUSTech-Competition, instead of using a subset of CASIA-E. Additionally, the training set was not provided, and participants were required to collect their own training data. This change introduced a cross-domain challenge, making the competition more demanding compared to previous editions. Our aim is to encourage the research community to develop gait recognition methods that are suitable for a wider range of applications. Despite the increased difficulty, participants in HID 2023 demonstrated their exceptional capabilities and achieved promising results.

The paper is structured as follows. In Section 2, we provide an overview of the competition, including details about the dataset, evaluation metric, fair competition organization, and statistical information. Section 3 presents the results achieved by the top-performing teams, along with descriptions of their methods. Some analyses are given in Section 4. Finally, in Section 5, we conclude the paper.

2. The Competition

2.1. Dataset

Since the accuracy in the previous competition is approaching saturation, we have introduced a new dataset, SUSTech-Competition, for HID 2023. The dataset was collected during the summer of 2022, with the approval of the Southern University of Science and Technology Institutional Review Board. It comprises 859 subjects and encompasses various variations, including clothing, carrying conditions, and view angles, as depicted in Figure 2. To alleviate the participants' data preprocessing burden, we provided human body silhouettes in the competition. These silhouettes were obtained from the original videos using a deep person detector and a segmentation model provided by our sponsor, Watrix Technology.

Figure 2. Some RGB images from the dataset SUSTech-Competition. Many variations are included in the dataset.

All silhouette images were resized to a fixed size of 128×128, as illustrated in Figure 3. We intentionally did not manually remove low-quality silhouettes, as the presence of noise reflects real-world application scenarios and adds to the challenge of the competition. This approach ensures that the competition provides a realistic simulation of real applications.

Figure 3. Some silhouettes from the dataset SUSTech-Competition.

Unlike the previous competition, we did not provide a specific training set to participants. Instead, participants were given the freedom to use any dataset, such as CASIA-B [17], OUMVLP [13], CASIA-E [12], GREW [20], Gait3D [18], or their own datasets, to train their algorithms. The cross-domain challenge was introduced to encourage participants to consider this aspect for achieving optimal results. The gallery set consists of only one sequence per subject, with the labels of the sequences provided to the participants. On the other hand, the probe set contains five randomly selected sequences per subject. The probe samples may exhibit variations in view, clothing, carrying conditions, and occlusions compared to the gallery samples. These settings make the competition more challenging and align it more closely with real-world applications compared to the previous three editions.

2.2. Performance metric

The evaluation metric used in HID 2023, as in previous competitions, is rank-1 accuracy, which provides a straightforward and easily implemented metric. It can be calculated as follows:

$$\text{Accuracy} = \frac{TP}{N} \tag{1}$$

where TP represents the number of true positives, and N corresponds to the total number of probe samples.

2.3. Competition organization

The evaluation process for HID 2023 was designed to be user-friendly, convenient, and secure against hacking attempts. The following rules were established to meet these requirements:

1. The competition consists of two phases. The first phase runs from February 15 to April 5, 2023, with only 10% of the test samples. The second phase takes place from April 6 to April 15, 2023, and includes the remaining 90% of the samples. The results obtained in the second phase are considered final. The first phase is 45 days long, while the second phase is significantly shorter, with only 10 days. This design was implemented to prevent sample label hacking.

2. To prevent the ID labels of the probe set from being deduced through multiple submissions, each team is limited to a maximum of 5 submissions per day during the first phase and 2 submissions per day during the second phase. Only one CodaLab ID is allowed per team, and only registrations using institutional emails (not public emails) are accepted.

3. The accuracy of the submissions is automatically evaluated on CodaLab, and the rankings are updated on the scoreboard accordingly. This immediate feedback ensures a user-friendly evaluation process.

4. The top teams on the final scoreboard are required to submit their programs to the organizers. The submitted programs are executed to reproduce their results, and the reproduced results should align with those displayed on the CodaLab scoreboard.

A total of 323 registrations were received for HID 2023, and registrations with public emails (e.g., Gmail) were rejected. Among the valid registrations, which amounted to 254, 159 teams submitted their results to CodaLab during the second phase. The programs of the top teams were carefully evaluated to verify the reproducibility of their results. After a thorough evaluation, the top 8 teams were selected based on their performance. The methods employed by these top teams will be discussed in the following section.

3. Methodologies

The organizers extended an invitation to the top teams to submit their source code for review. Out of the 8 teams, all of them submitted their source code, while 7 teams also provided detailed method descriptions. The technologies utilized by these teams, along with their corresponding results, are summarized in Table 1. The subsequent part of the section provides an in-depth exploration of the methodologies employed by each team.

3.1. Team: Terrance

Member: Chenye Wang (School of Mathematics and Statistics, Shandong University) {201900810349@mail.sdu.edu.cn}

Method: In the proposed method, the GaitBase model from OpenGait [4] is utilized as the training framework, with ResNet18 serving as the backbone. To augment the data, horizontal flipping and random erasing are employed. The training datasets consist of the CAISA-E dataset [12] and the high-quality CAISA-B* dataset [8]. The data preprocessing pipeline comprises four steps. Firstly, silhouette sequences containing completely black or white images are eliminated by examining the foreground region pixel ratios. Secondly, the midpoints of the upper and lower body are detected and then connected to rotate the image, aligning it perpendicular to the x-axis. Thirdly, the alignment is performed using the same technique as OpenGait. Finally, all training silhouettes are resized to a fixed size of 128×128 pixels.

To enhance the model's cross-domain capability, the self-training mechanism is incorporated during the training process [7]. Specifically, multiple pre-trained models are employed for testing and identifying samples that meet certain conditions, such as high confidence, to be considered as high-confidence samples. Then the predicted labels of these high-confidence samples are utilized as pseudo-labels. The two conditions are: (1) The dissimilarity score of the probe and gallery is small, and (2) The dissimilarity score difference between rank-1 and rank-2 is large.

In the final stage of the method, the model underwent fine-tuning using high-confidence samples and their corresponding pseudo-labels, which demonstrated significant

Table 1. The technologies used by the top 8 teams and their accuracies in HID 2023. GaitBase* indicates GaitBase model without residual learning, and GaitBase*$^{+}$ denotes GaitBase* model with Batch normalization layer (BN) on the last two layers.

Team rank	1	2	3	4	5	6	7	8
CodaLab ID	Terrance	fried-chicken	league	GRgroup	Li_qing	cartes1us	WBH	Zekai
Data cleaning	✓	×	×	×	×	×	×	✓
Data alignment	✓	✓	✓	✓	✓	×	✓	✓
Data augmentation	✓	✓	✓	✓	✓	✓	✓	✓
Query expansion	×	✓	✓	✓	✓	×	✓	×
Re-ranking	×	✓	✓	✓	✓	×	✓	×
Ensemble	×	✓	✓	✓	×	✓	✓	✓
Training data	CASIA-E, CASIA-B	HID2022	Grew, HID2022	CASIA-E	HID2022	Gait3D, HID2022	HID2022	Gait3D, CASIA-E
Pseudo-labelling	✓	✓	×	×	×	×	×	×
Architecture	GaitBase [4]	VideoResnet [14]	DeepGaitV2-P3D [3]	GaitBase* [4], GaitMask [9], GaitGL [10]	GaitBase* [4]	GaitBase* [4], GaitBase [4], DeepGaitV2-P3D [3]	GaitBase* [4], GaitBase*$^{+}$	GaitBase [4]
GPU	RTX3090*8	A100*4	RTX3090*2	V100*8	RTX2080*2	A6000*4	RTX3090*2	N/A
Accuracy(%)	80.8	75.5	73.0	64.5	61.6	58.3	57.4	57.1

cross-domain potential. To achieve this, the original training set is combined with high-confidence samples in a 2:8 ratio, and a new training set is created. By exclusively employing triplet loss to train the last two layers of the backbone, the model's performance improved approximately by 8%. The results show that the self-training mechanism can further enhance both the quality and quantity of pseudo-labels, thereby leading to additional improvements in the model's performance.

The experimental settings were as follows: Eight GeForce RTX 3090 GPUs were for training. The batch size was set to 32×8. There were 65,000 iterations in the pre-training stage and 8,000 iterations in the fine-tuning stage. The remaining hyperparameters were mostly kept consistent with those of the GaitBase model in OpenGait.

3.2. Team: fried-chicken

Supervisor: Annan Li (Beihang University) {liannan@buaa.edu.cn}
Member: Yuwei Zhao (Beihang University) {sy2206328@buaa.edu.cn}

Method: This method is built upon the OpenGait [4] framework and encompasses 5 aspects: *(1) 3D Model:* A 3D backbone is implemented using VideoResnet [14] from torchvision, with the residual blocks structure set as $[1, 4, 4, 1]$. *(2) Data Augmentation:* The default data augmentation strategy from GaitBase is employed in OpenGait, with the exception of adjusting the probability of horizontal flip from 0.2 to 0.5. *(3) Ensemble:* By following the method employed by team *league* in HID 2022 [16], two models are trained and the second model is trained using the flipped version of the entire sequences. To create an ensemble, the average Euclidean distance is calculated between the embeddings of the two models. *(4) Pseudo Label:* Initially, the model was trained using the training data (500 people) from HID 2022 and then utilized this model to generate pseudo labels for the testing data (504 people) of HID 2022. All the data (1004 people) was utilized to retrain the model. *(5) Re-ranking:* HID 2023 consists of two distinct phases, with non-overlapping test data. It was observed that re-ranking [19] during the first phase led to a degradation in performance, while re-ranking during the second phase resulted in performance gains. This insight highlighted the significant impact of the number of test samples on re-ranking performance. To leverage the data from the first phase fully, it was included for re-ranking even during the second phase.

Due to the limited number of daily submissions, a thorough ablation study could not be conducted. However, it was roughly estimated that the performance gains from the five aforementioned methods were approximately +6%, +1%, +3%, +4%, and +1.5%, respectively.

3.3. Team: league

Member: Li Wang and Lichen Song (Dalian Everspry Sci&Tech Co., Ltd.) {challenge@everspry.com}

Method: The employed model, DeepGaitV2-P3D, adopts an ensemble approach utilizing two backbones. The backbones consist of a 10-layer ResNet and a 22-layer ResNet, respectively. The structure of the 3D convolution module is derived from the P3D [11] architecture. The overall model architecture is inspired by DeepGaitV2 in [3]. Data augmentation techniques employed include random perspective transformation, random horizontal flipping, and random rotation. To enhance the performance, features extracted from both the original and flipped samples are fused, and the features obtained from the 10-layer backbone are combined with those from the 22-layer backbone.

Prior to training on the HID 2023 dataset, DeepGaitV2-P3D is pre-trained on the GREW dataset [20]. Subsequently, it undergoes fine-tuning on the gallery set of the HID 2023 test dataset and further fine-tuning on the HID 2022 dataset. Finally, the re-ranking strategy is employed to refine the recognition accuracy.

3.4. Team: GRgroup

Supervisor: Shunli Zhang, Xiang Wei, Jiande Sun, and Yang Yang.
Member: Ming Wang, Qianying Tang, Senmao Tian, Junzhe Chen (Beijing Jiaotong University) {21121736@bjtu.edu.cn}
Method: The overview of the proposed method is shown in Figure 4. The CASIA-E [12] dataset was used for training. The method includes three parts: data augmentation, network design, and test strategies. For training, the data augmentation includes a random horizontal flip and a random rotation. Three different networks were trained: baseline[4], GaitMask [9], and GaitGL [10]. For the test phase, gait features were firstly extracted, and then QE (Query Expansion) [15] and RK (re-Rank King) [19] were employed to improve the retrieval accuracy. Finally, the VM (Vote Mechanism) was used to improve recognition accuracy. The results of the three models and their ensemble are shown in Table 2. It shows that the ensemble can improve the accuracy obviously.

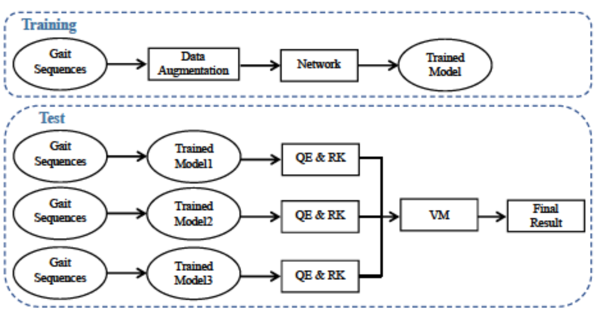

Figure 4. The framework of Team GRgroup's method.

Table 2. Rank-1 accuracy (%) of the Baseline, GaitMask, GaitGL, and their ensemble respectively for Team GRgroup.

Model	Accuracy
Baseline	60.1%
GaitMak	61.1%
GaitGL	62.5%
Ensemble	64.5%

3.5. Team: Li_qing

Member: Qing Li and Xianqiang Yang (Harbin Institute of Technology) {1190100228@stu.hit.edu.cn}
Method: The primary focus of the method lies in data preprocessing and augmentation. The training dataset from HID 2022 [16] was utilized. The data preprocessing steps followed the same approach as OpenGait [4]. Given that both the training and test datasets contained noise, the low-quality data was not cleaned from the training set and instead used all original images. The data augmentation process involved random horizontal flipping and random angle rotation. Additionally, the re-ranking technique was incorporated to enhance performance.

During the training phase, two Geforce RTX 2080 GPUs were uesed. The input image width was standardized to 128 pixels. SGD was employed as the optimizer, with a base learning rate of 0.1. The training batch size was set to [8, 4], and the total number of training iterations amounted to 120K. The optimizer scheduler gamma was 0.1, resulting in a learning rate reduction at 20K, 40K, and 60K iterations. Other settings remained consistent with the default configurations of the OpenGait baseline. Ultimately, the weights from iteration 110K were chosen.

3.6. Team: cartes1us

Member: Wenlong Li, Likai Wang, Yuchao Zhong, and Jingyu Zhang (Xidian University) {wenlongli, lkone, yuchaozhong, jingyuzhang}@stu.xidian.edu.cn
Method: During the training phase, two backbones were trained: the baseline model in OpenGait [4] and the proposed GaitBase-P3D model. To create a stronger gait recognition model through ensemble learning, distance normalization techniques were integrated from these models. Inspired by [3, 11], the 2D convolution in GaitBase [4] was replaced with Pseudo 3D convolution to construct the Pseudo 3D convolution ResNet-like backbone. The Pseudo 3D convolution offers reduced computational cost compared to the standard 3D convolution. To eliminate the impact of dimension, the distance between a probe sample and a gallery sample was normalized for each model. Then, gait recognition is performed by computing a weighted sum of the scores obtained from the two base models. The pipeline of the ensemble learning approach is illustrated in Figure 5.

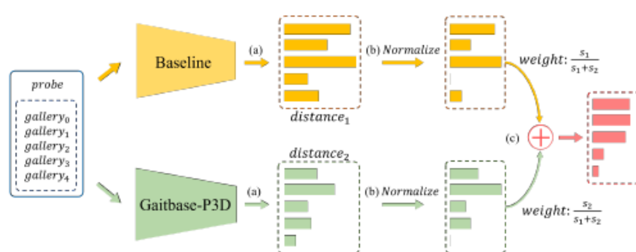

Figure 5. The pipeline of Ensemble Learning from Team cartes1us.

The data augmentation probability for horizontal flip was set to 0.5, while the details of rotation and perspective transformation followed the approach in [4]. Techniques such as data cleaning, data alignment, query expansion, and re-ranking were not employed. The GaitBase-P3D structure and other training settings were similar to the 22-layer DeepGaitV2-3D [3], with the number of basic feature map channels set to 64. The overall experiment was built upon the OpenGait framework, utilizing four NVIDIA A6000 GPUs. The experimental results are presented in Table 3. Ensemble learning different modeling methods and training data is more conducive to improving the performance of ensemble learning.

	Model	Training data	Input	Accuracy
1	Baseline	HID 2022	128×88	49.0%
2	GaitBase	Gait3D training	64×44	50.7%
3	GaitBase-P3D	Gait3D training	64×44	51.8%
4	GaitBase-P3D	Gait3D all	64×44	54.7%
	Ensemble(1,3)	-	-	56.2%
	Ensemble(1,4)	-	-	58.3%

Table 3. Experimental results of Team cartes1us: Training GaitBase-P3D with all data in the Gait3D dataset can improve generalization.

3.7. Team: WBH

Member: Runsheng Wang, Zongyi Li, Bohao Wei, He Li, Jianbo Li, Hefei Ling, and Yuxuan Shi. (Huazhong University of Science and Technology) {wrsh, zongyili, xavid, he_li, m202273875, lhefei, shiyx}@hust.edu.cn

Method: For ensemble learning, two models were utilized: Baseline and BNBaseline. The Baseline model is from OpenGait [4], while BNBaseline is an extension of the Baseline with the addition of Batch Normalization (BN) layers. The integration of BN layers is motivated by their empirical effectiveness in domain transfer tasks [1]. In the approach, BN layers are applied to the last two layers of the network. During training, the triplet loss and the classification loss with BNNeck were employed, following the practices of most state-of-the-art methods. Additionally, two data augmentation strategies, namely random rotation and random erasing, were utilized.

To combine the outputs of Baseline and BNBaseline, a weighted ensemble strategy was employed. This involved assigning weights to the Euclidean distances computed using the features generated by the two models. Since BNBaseline outperforms Baseline, a higher weight was assigned to the former. Finally, the K-reciprocal re-ranking strategy [19] to refine the retrieval results.

Ablation studies were conducted to demonstrate the effectiveness of various strategies. The results of these ablation studies are presented in Table 4. The findings highlight the efficacy of BN layers in the cross-domain gait recognition task. Additionally, the weighted ensemble and re-ranking strategies exhibit a significant performance boost, as indicated by the noticeable improvements observed.

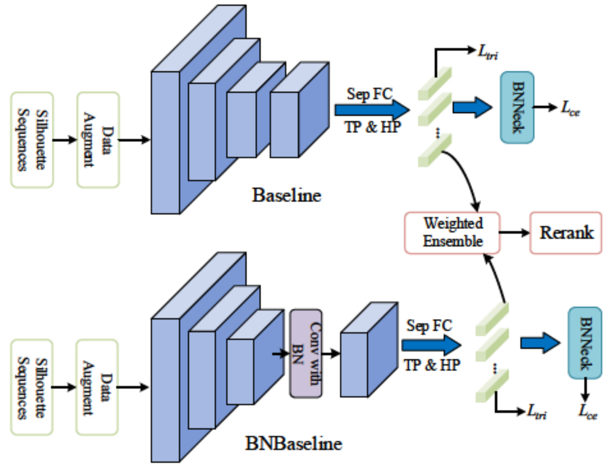

Figure 6. The framework of team WBH's solution: Two models, Baseline and BNBaseline, were employed for ensemble learning. (SepFC is separate fully connected layers, TP is temporal pooling, and HP is horizontal pooling)

Table 4. Ablation study of different models and options by team WBH.

Baseline	BNBaseline	Re-ranking	Accuracy
✓			49.3%
	✓		51.3%
	✓	✓	51.7%
✓	✓		53.0%
✓	✓	✓	57.4%

4. Analysis

According to Table 1 and the method descriptions provided by the top teams, several important technologies for gait recognition can be identified. To evaluate the contributions of these technologies, some ablation experiments are needed. Since the experiments were carried out by different teams, fair ablation experiments are not easy to implement. We just list the technologies here.

- **Data cleaning**: Only 2 out of the top 8 teams utilized data cleaning techniques, in contrast to the 5 out of 10 teams that employed it in the previous competition. Although noisy data may improve the generalization capability of trained models, the absence of training data in this cross-domain challenge may have influenced the reduced usage of data cleaning methods.

- **Data alignment**: Data alignment, either in the spatial or temporal domain, proved to be beneficial. 7 out of the 8 top teams incorporated data alignment techniques, while only 2 out of the top 10 teams in HID 2022 utilized it. Temporal domain alignment is particularly challenging to implement compared to spatial domain alignment.

- **Data augmentation**: All top teams employed data augmentation, which has become a standard preprocessing step in deep learning tasks. Data augmentation enriches the training samples and enhances the robustness of trained models against various variations.

- **Query expansion**: Query expansion [6], a technique that combines highly ranked samples from an original query into an expanded query, was employed by 5 out of the top 8 teams, compared to 4 out of the 10 teams in the previous competition. Query expansion has been shown to improve accuracy in gait recognition.

- **Re-ranking**: Similar to HID 2021 [15] and HID 2022x [16], almost all teams incorporated re-ranking into their methodologies. Re-ranking, as described in [19], significantly enhances accuracy. 6 out of the top 8 teams employed re-ranking in their experiments, although it introduces additional computational costs.

- **Pseudo-labelling**: Pseudo-labelling was utilized by the top 2 teams. Team 1 (Terrance) achieved an accuracy of 80.8% without re-ranking or ensemble learning, demonstrating a substantial lead over Team 2. The top 2 teams employed pseudo-labelling, but there was a difference in their implementations. Team 1 fine-tuned the model using the probe samples in the test set, while Team 2 did not. While some may argue the fairness of using test data for model fine-tuning, it is important to note that no labels were employed during the fine-tuning process.

- **Backbone model**: The backbone models play a crucial role in gait recognition as they are responsible for extracting discriminative features. In the competition, all teams relied on well-tested models from the literature or made modifications to existing models. None of the teams proposed completely new backbone models for the competition.

In HID 2023, the highest accuracy achieved was 80.0%, whereas in HID 2022, it reached 95.9%. Upon analyzing the datasets from both competitions, we have identified three primary factors contributing to these differences. (1) The first is the cross-domain challenge. It played a significant role. HID 2023 did not provide training data. The training data that the participants used has different scenarios from the test data. The gap also indicates that the current state-of-the-art methods still have no good generalization capability. This limitation highlights the need for further improvements in achieving robustness across different domains. (2) Secondly, there was an increase in the number of subjects within the test set. In HID 2020-2022, the test set comprised 505 subjects, whereas in HID 2023, it expanded to 859 subjects. This increase in the number of subjects inherently leads to a decrease in accuracy, even when utilizing the same methods. (3) HID 2023 introduced a greater variety of human activities into the SUSTech-Competition dataset. This diversification of activities makes the competition more challenging.

We believe these challenges have undoubtedly pushed the boundaries of research and inspired further advancements in the field.

5. Conclusion

Competitions serve as valuable indicators of progress, and this particular competition highlights the effectiveness of gait recognition in increasingly challenging scenarios. Over the four competitions, gait recognition demonstrated its potential as a reliable biometric modality. This competition has particularly stood out, reaching new heights of achievement despite introducing a considerably more challenging setting compared to its predecessors. This remarkable progress underscores the continuous improvement and innovation within the field of gait recognition, setting a new benchmark.

Acknowledgements

We express our sincere appreciation to Watrix Technology Limited Co. Ltd. for their generous sponsorship of the competition, which played a crucial role in its success. We would also like to express our gratitude to CodaLab for providing the result evaluation platform, enabling a seamless and efficient evaluation process. We are deeply grateful to our advisors, Prof. Mark Nixon, Prof. Tieniu Tan, and Prof. Yasushi Yagi, for their invaluable support and insightful suggestions throughout the competition. We would also like to acknowledge the technical support provided by Mr. Jingzhe Ma, Mr. Rui Wang, and Mr. Junhao Liang. Their assistance and expertise greatly contributed to the successful organization and execution of the competition.

References

[1] W.-G. Chang, T. You, S. Seo, S. Kwak, and B. Han. Domain-specific batch normalization for unsupervised domain adaptation. In *CVPR*, pages 7346–7354, 2019. 6

[2] H. Chao, K. Wang, Y. He, J. Zhang, and J. Feng. Gaitset: Cross-view gait recognition through utilizing gait as a deep set. *IEEE TPAMI*, 44(7):3467–3478, 2022. 1

[3] C. Fan, S. Hou, Y. Huang, and S. Yu. Exploring deep models for practical gait recognition. *arXiv preprint arXiv:2303.03301*, 2023. 4, 5

[4] C. Fan, J. Liang, C. Shen, S. Hou, Y. Huang, and S. Yu. OpenGait: Revisiting gait recognition toward better practicality. In *CVPR*, 2023. 3, 4, 5, 6

[5] C. Fan, Y. Peng, C. Cao, X. Liu, S. Hou, J. Chi, Y. Huang, Q. Li, and Z. He. GaitPart: Temporal part-based model for gait recognition. In *CVPR*, pages 14213–14221, 2020. 1

[6] A. Gordo, F. Radenovic, and T. Berg. Attention-based query expansion learning. In *ECCV*, pages 14213–14221, 2020. 7

[7] D.-H. Lee. Pseudo-Label: The simple and efficient semi-supervised learning method for deep neural networks. In *ICML Workshop on Challenges in Representation Learning*, 2013. 3

[8] J. Liang, C. Fan, S. Hou, C. Shen, Y. Huang, and S. Yu. Gait-Edge: Beyond plain end-to-end gait recognition for better practicality. In *European Conference on Computer Vision*, pages 375–390, 2022. 3

[9] B. Lin, X. Yu, and S. Zhang. Gaitmask: Mask-based model for gait recognition. In *BMVC*, pages 1–12, 2021. 4, 5

[10] B. Lin, S. Zhang, and X. Yu. Gait recognition via effective global-local feature representation and local temporal aggregation. In *ICCV*, pages 14648–14656, 2021. 1, 4, 5

[11] Z. Qiu, T. Yao, and T. Mei. Learning spatio-temporal representation with pseudo-3d residual networks. In *ICCV*, pages 5534–5542, 2017. 4, 5

[12] C. Song, Y. Huang, W. Wang, and L. Wang. CASIA-E: A large comprehensive dataset for gait recognition. *IEEE TPAMI*, 45(3):2801–2815, 2023. 2, 3, 5

[13] N. Takemura, Y. Makihara, D. Muramatsu, T. Echigo, and Y. Yagi. Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition. *IPSJ Trans. on Computer Vision and Applications*, 10(4):1–14, 2018. 2

[14] D. Tran, H. Wang, L. Torresani, J. Ray, Y. LeCun, and M. Paluri. A closer look at spatiotemporal convolutions for action recognition. In *CVPR*, pages 6450–6459, 2018. 4

[15] S. Yu, Y. Huang, L. Wang, Y. Makihara, E. B. Garcia Reyes, F. Zheng, M. A. R. Ahad, B. Lin, Y. Yang, H. Xiong, B. Huang, and Y. Zhang. HID 2021: Competition on human identification at a distance 2021. In *2021 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–7, 2021. 5, 7

[16] S. Yu, Y. Huang, L. Wang, Y. Makihara, S. Wang, M. A. Rahman Ahad, and M. Nixon. HID 2022: The 3rd international competition on human identification at a distance. In *2022 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–9, 2022. 2, 4, 5, 7

[17] S. Yu, D. Tan, and T. Tan. A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In *ICPR)*, volume 4, pages 441–444. IEEE, 2006. 1, 2

[18] J. Zheng, X. Liu, L. H. Wu Liu, C. Yan, and T. Mei. Gait recognition in the wild with dense 3d representations and a benchmark. In *CVPR*, 2022. 1, 2

[19] Z. Zhong, L. Zheng, D. Cao, and S. Li. Re-ranking person re-identification with k-reciprocal encoding. In *CVPR*, pages 3652–3661, 2017. 4, 5, 6, 7

[20] Z. Zhu, X. Guo, T. Yang, J. Huang, J. Deng, G. Huang, D. Du, J. Lu, and J. Zhou. Gait recognition in the wild: A benchmark. In *ICCV*, 2021. 1, 2, 4