Human Identification at a Distance: Challenges, Methods and Results on the Competition HID 2024

Shiqi Yu[1], Weiming Wu[2], Jiacong Hu[2], Zepeng Wang[3], Jingjie Wang[4], Meng Zhang[5], Runsheng Wang[6], Yunfei Ni[6], Yongzhen Huang[7,8], Liang Wang[9], and Md Atiqur Rahman Ahad[10]

[1]Southern University of Science and Technology, China. [2]South China University of Technology, China.

[3]Shanghai Jiao Tong University, China. [4]Beijing Jiaotong University, China. [5]Fujitsu Limited, Japan.

[6]Huazhong University of Science and Technology, China. [7]Beijing Normal University, China.

[8]Watrix Technology Limited Co. Ltd. [9]Institute of Automation, Chinese Academy of Sciences, China.

[10]University of East London, UK.

https://hid.iapr-tc4.org/

Abstract

Human identification at a distance (HID) faces challenges due to the difficulty of acquiring traditional biometric modalities like face and fingerprints. Gait recognition offers a viable solution since it can be captured at a distance. To advance the algorithm development and provide fair evaluations, the International Competition on Human Identification at a Distance (HID) has been held annually since 2020, with HID 2024 marking the fifth edition. Despite increased difficulty, participants demonstrated remarkable capabilities, surpassing previous accuracy levels. This paper, co-authored by competition organizers and top participants, provides a comprehensive summary of HID 2024, including an overview of the competition, and insights into the methods employed by the top teams. Specifically, inspired by the achievements of the 5 competitions of HID, we also provide the insights for the future directions on gait recognition.

1. Introduction

Human identification at a distance faces many challenges [11] because most traditional biometric modalities, like faces, fingerprints, iris, are difficult to acquire. Gait may be the only promising biometric modality [9, 2] for this purpose because it can be collected even when faces are obscured or too small to be detected.

In the past two decades, gait recognition has been improved obviously particularly with deep learning. Some typical algorithms, such as GaitSet [1], GaitGL [7] and GaitBase [4], have been developed, showing promising results. Some recent methods like BigGait [15] employs large vision models to improve gait recognition. However, the accuracy 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 [21, 19]. The studies in [4] show the gait recognition models trained on some indoor datasets cannot achieve promising results on in-the-wild datasets. 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. HID 2024 is the 5th edition of the series. In the first 3 competitions, the dataset was a subset of CASIA-E [13]. The accuracy on it reached to 95.9% in HID 2022, and was almost saturated. So a new dataset, *SUSTech-Competition*, was introduced in HID 2023 for the first time. The full set of CASIA-E was made public available after its introduction paper [13] was accepted by IEEE TPAMI. Since HID 2023, there is no training data provided by the competition organizers. Participants needs to collect their own training data. This change can introduce a cross-domain challenge. It makes the competition more challenging compared to previous editions. Our aim is to encourage the research community to develop gait recognition methods for a wider range of applications. Despite the increased difficulty, the participants in HID 2023 demonstrated their exceptional capabilities and achieved promising results. The participants in HID 2024 put the accuracy to a higher level than the previous HID 2023 as show in Figure 1.

The paper has been composed by the competition organizers and the participants from the top teams. It summaries the competition HID 2024. Specifically, in Section 2, an overview of the competition is provided. It includes details

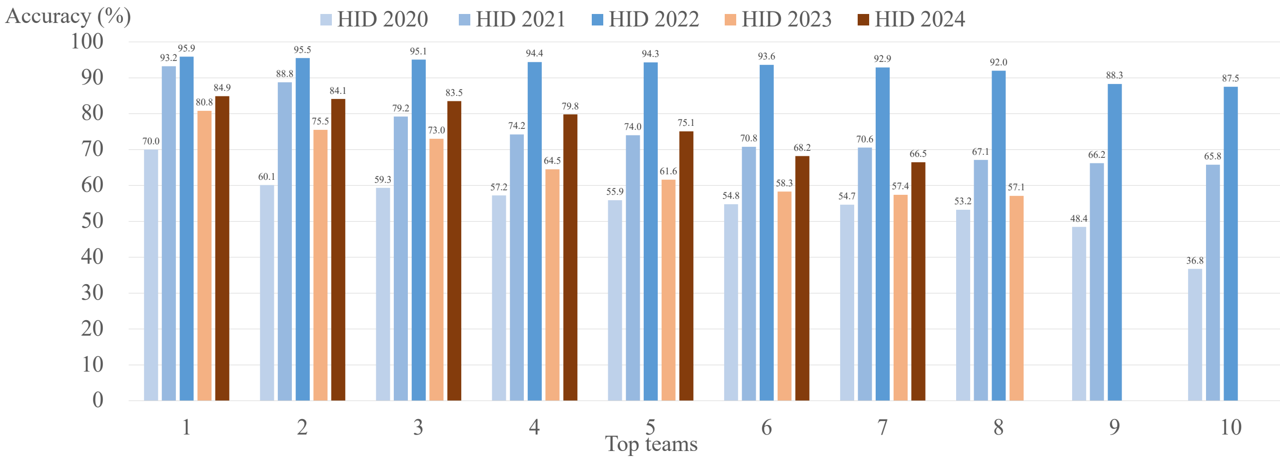

Figure 1. The top results of HID 2024 and the previous 4 competitions. The dataset for HID 2020, 2021 and 2022 is CASIA-E, and it is *SUSTech-Competition* for HID 2023 and 2024. The results of HID 2020 and HID 2021 have been calibrated according to the same standard as HID 2022.

about the dataset, the evaluation metric, fair competition organization, and some 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. Dataset and Competition Details

2.1. Dataset

The dataset used in HID 2024 is the same as the one in the previous competition and is *SUSTech-Competition*. However, the test set has been changed by randomly selecting another set of samples. The dataset was collected during the summer of 2022, with the approval of the Southern University of Science and Technology Institutional Review Board. The full dataset comprises 859 subjects and encompasses various variations, including clothing, carrying conditions, and view angles, as shown 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.

All silhouette images were resized to a fixed size of 128×128, as illustrated in Figure 2. 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.

The same as HID 2023, we did not provide a specific training set to participants. Instead, participants can use any dataset, such as CASIA-B [17], OUMVLP [14], CASIA-E [13], GREW [21], Gait3D [19], SUSTech-1K [12], CCPG [6], CCGR [22], DroneGait [5] or their own datasets, to train their models. Since the training set and the test set would be from different datasets, the cross-domain challenge will be introduced. Participants have to consider this aspect for achieving good results. The gallery in the test set consists of only one sequence per subject, with the labels of the sequences provided to the participants. The probe 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 challenging and align it closely with real applications.

Figure 2. Some RGB images and their corresponding silhouettes from the dataset *SUSTech-Competition*. Many variations are included in the dataset.

Specifically, the test set in HID 2024 is different from the one in HID 2023 even they are both from *SUSTech-Competition*. They are all randomly selected subsets from *SUSTech-Competition*. Besides, the samples from the test probe set, including HID 2023 and all the test probe set samples provided in HID 2024, have been strictly prohibited from being used in any way during the training phase.

2.2. Performance metric

As the previous competitions of the series, the evaluation metric is the 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 policies

The evaluation process for HID 2024 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 March 11 to May 10, 2024, with only 10% of the test samples. The second phase takes place from May 11 to May 20, 2024, and includes the remaining 90% of the samples. The results obtained in the second phase are considered final. The first phase is 2 months 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.

2.4. Competition statistics

A total of 90 registrations were received for HID 2024, and registrations with public emails (e.g., Gmail) had been rejected. Among the valid registrations, which amounted to 23 teams submitted their results to CodaLab during the second phase. The best scores and the numbers of submissions of each day can be found in Figure 3. Most teams actively participated the competition. The programs of the top teams were carefully evaluated to verify the reproducibility of their results. After a thorough evaluation, the top 7 teams were selected based on their performance. The methods employed by these top teams will be introduced in the following section.

3. Methods of the Top Teams

The competition organizers invited all top teams to submit their source code for review. Seven teams submitted their course code, and passed the code review. The 7 teams also provided their method descriptions. The subsequent part of the section provides an in-depth exploration of the methods employed by each team. The technologies utilized by these teams, along with their corresponding results, are summarized in Table 1.

3.1. Team: SCUT-BIPLAB

Members: Weiming Wu, and Kun Liu
South China University of Technology
{auauweimingwu, aulkun}@mail.scut.edu.cn
Method: The method utilized the OpenGait [4] framework for model training, specifically using the DeepGaitV2 series, which includes the P3D and 3D modules, both underpinned by 22-layer ResNet architectures. The training regimen incorporated the framework's standard data augmentation algorithms, employing a batch size of 16×8 over 130,000 iterations. Hyperparameters are those predefined for the DeepGaitV2 model in OpenGait. The experiment were on 2 pieces of RTX 3090 GPUs.

Dataset merging was crucial for model training, and can involve multiple scenarios such as normal walking, carrying objects like umbrellas and backpacks, and different clothes including pants and skirts. The merged dataset consisted of silhouette sequences from several popular datasets, Gait3D, CASIA-B, CCPG, and SUSTech1K. Some selected sequences from CASIA-E, OUMVLP, and CCGR were also included. In total there were 8,940 subjects across diverse viewing angles in the training set. Each training sample was resized to 64.

The test samples exhibited a range of challenges, including low quality, noise and distortions. A random sampling approach was employed to eliminate low-quality samples and noise. Some rotations were also introduced to data augmentation. These preprocessing steps were applied to both the gallery and probe samples.

Post-training, the models, particularly the P3D and 3D variants, were deployed to analyze the gallery and probe test samples. Final model outputs were fused through an ensemble approach leveraging multiple inference results. This ensemble was executed by conducting dual inferences per model: the first maintaining the original sequence order of samples, and the second processing them in reverse. The ultimate results submitted were derived by calculating the

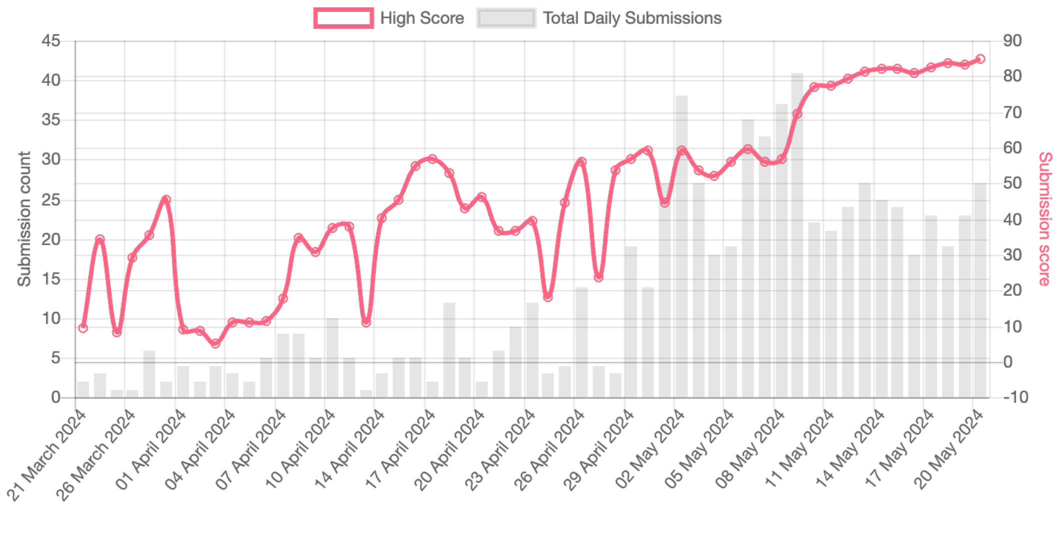

Figure 3. The best scores and the numbers of submissions of each day during the competition.

Table 1. The technologies used by the top 7 teams and their accuracies in HID 2024.

Team rank	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Team Name	SCUT-BIPLAB	jchu	SJTU-ICL	GRgroup	BRAVO-FJ	dashengge	HUST-MCLAB
Data cleaning	✓	×	×	✓	×	×	×
Data alignment	✓	✓	✓	✓	✓	✓	✓
Data augmentation	✓	✓	✓	✓	✓	✓	✓
Re-ranking	✓	✓	✓	✓	✓	✓	✓
Ensemble	✓	✓	✓	×	✓	×	✓
Training data	Gait3D, CASIA-B, CCPG, SUSTech-1K, CASIA-E, OUMVLP, CCGR	Gait3D, CASIA-B, CCPG, SUSTech-1K	CCPG, Gait3D, GREW, CASIA-B, CASIA-E, OUMVLP, SUSTech-1K	CAISA-E	GREW, HID 2022	HID 2022	SUSTech-1K, Gait3D
Pseudo-labelling	×	×	×	×	✓ (on HID2022 data)	×	×
Architecture	DeepGaitV2 (P3D&3D) [3]	DeepGaitV2 (3D) [3]	DeepGaitV2 (P3D) [3], SwinGait [3]	GaitGL+Gem [7]	DeepGaitV2 (P3D) [3]	DeepGaitV2 (P3D), SwinGait, GaitBase [4]	DeepGaitV2 (P3D), SwinGait [3]
GPU	RTX 3090 * 2	N/A	A100 * 4	RTX 3090 * 4	A6000*4	N/A	N/A
Accuracy(%)	84.9	84.1	83.5	79.8	75.1	68.2	66.5

mode from these multiple inference sequences, ensuring robust and accurate outcomes.

3.2. Team: jchu

Members: Jiacong Hu, and Xiaochuan Liao
South China University of Technology
{202321018015, au_tairistu}@mail.scut.edu.cn

Method: The DeepGaitV2-3D model is employed to tackle the challenges posed by diverse gait data. The whole process includes dataset selection, data augmentation, network design, and ensemble learning. To improve the model's generalization capabilities, a comprehensive training set was compiled from four datasets: Gait3D, CASIA-B, CCPG, and SUSTech1K. The 4 datasets contains both indoor and outdoor data, as well as constrained and unconstrained environments. Data augmentation includes rotation, horizontal flipping, perspective changes, cropping, etc. A novel temporal technique termed reranking, which reconstructs sequences by truncating at random points and reversing the order of spliced subsequences. To further improve the model, a modified Hierarchical Progressive Pooling (HPP) architecture is employed to replace the standard GAP+GMP with a summation of the top two maxima and integrating a two-layer MLP-like attention module prior to the BNNeck. This method, referred to as DGv2+, adjusts weights across different sites and channels of embeddings.

Ensemble learning was leveraged by three distinct traing configurations: a 22-layer DGv2+_22 model with unaligned data, a 22-layer model with spatially aligned data, and a 30-layer model also with unaligned data. These models were trained with a batch size of 8×8, 240,000 iterations for the 22-layer models and 260,000 for the 30-layer model. Ensemble decision-making was implemented by aggregating votes from the three models to determine the most possible ID of a given gait sequence. Individual and integrated performances of these models are summarized in Table 2.

Table 2. The rank-1 accuracies of three models and their ensemble.

Method	Accuracy (%)
DGv2+_22 w/o alignment	82.10
DGv2+_22 w/ alignment	82.05
DGv2+_30 w/o alignment	82.90
Ensemble	84.14

3.3. Team: SJTU-ICL

Members: Zepeng Wang, and Ke Xu
Shanghai Jiao Tong University
{wzp.ck, 113025816}@sjtu.edu.cn
Method: The method includes two models: the CNN-based DeepGaitV2, and the Transformer-based SwinGait [3]. The models have been designed to adapt to the challenges of varied gait recognition data and incorporate strategies of dataset integration, advanced data augmentation, and novel network design enhancements. A comprehensive training set, by combining CASIA-E [13], CCPG [6], CASIA-B [17], Gait3D [19], SUSTech-1K [12], OUMVLP [14], and GREW [21], was formed to enhance the model's performance. The input size of images was fixed to 128×128 pixels.

In the test phase, the cosine similarity was chosen as the primary distance metric due to its enhanced performance in preliminary tests. Ensemble learning was adopted by training with 6 different models with various combinations of datasets and augmentation techniques. The final decision of a sequence was made through a voting mechanism among the 6 models for robust decision-making.

The implementation was carried out on OpenGait [4] with four A100 (40GB) GPUs. The initial training involved 360,000 iterations using an SGD optimizer with a detailed learning rate schedule that adjusted from 1×10^{-1} to 1×10^{-4}. The novel GaitMix augmentation was integrated at this stage to enhance the initial training. A fine-tuning was carried out using an AdamW optimizer to further improve the model performance over an additional 160,000 iterations with a learning rate schedule from 1×10^{-4} to 1×10^{-5}. Each training batch consisted of 32 identities with 4 samples for each.

3.4. Team: GRgroup

Members: Jingjie Wang, Jianlong Yu, Senmao Tian, and Ming Wang
Supervisor: Shunli Zhang and Xiang Wei
Beijing Jiaotong University
{23111492@bjtu.edu.cn}
Method: The method employs the GaitGL+Gem model as the primary training framework in OpenGait. Data augmentation techniques, specifically horizontal flipping and random erasing, are utilized to enhance dataset variability. The sole training dataset employed is the CAISA-E dataset.

The data preprocessing pipeline is executed in two phases. Initially, silhouette sequences that contain entirely black or white images are discarded after assessing the pixel ratios of the foreground regions. Subsequently, the midpoints of the upper and lower body in the remaining images are identified and used to rotate the silhouettes, ensuring alignment perpendicular to the x-axis.

To improve the model's cross-domain capability, it undergoes a fine-tuning phase using the HID2024 gallery dataset, which has demonstrated substantial potential for enhancing cross-domain applicability. The results from this phase indicate notable improvements in model performance.
Experimental Setup: The computational framework consisted of four GeForce RTX 3090 GPUs. The training regimen included a batch size of 32×8, encompassing 80,000 iterations during the pre-training phase and an additional 8,000 iterations during the fine-tuning phase. The rest of the hyperparameters were aligned with the standard configurations of the GaitGL model as specified in OpenGait.

3.5. Team: BRAVO-FJ

Member: Meng Zhang, and Kazuki Osamura
Supervisor: Rujie Liu, Narishige Abe, and Hidetsugu Uchida
Fujitsu Limited
{zhangmeng, osamura.kazuki, rjliu, abe.narishige, u.hidetsugu}@fujitsu.com

Method: The model employed in this research is DeepGaitV2 enhanced with a Pseudo 3D convolution module (P3D) [10]. Data augmentation techniques such as random perspective transformation, random horizontal flipping, and random rotation are applied, with respective probabilities of 0.3, 0.4, and 0.4. Both the softmax loss and the triplet loss, a margin of 0.2 for the latter, are utilized for training. The model undergoes initial pre-training on the GREW dataset [21] and subsequent fine-tuning on the HID2022 dataset [16].

A re-ranking strategy is implemented to improve performance. To leverage HID2022 data effectively, pseudo-labels for HID2022 test samples are generated through a voting mechanism involving several models, as depicted in Figure 4. If predictions for a HID2022 probe sample are consistent across multiple models, the sample will be used as a pseudo-label for further training; otherwise, it will be excluded. Various models with different backbones, specifically ResNet22, ResNet34, and ResNet38, are employed to construct the voting framework.

To enhance the reliability and stability of predictions, a voting mechanism is applied across different model iterations. After the final predictions for each model are obtained, this mechanism is engaged again to fuse the predictions from three different backbone models. It aims to integrate the strengths of each model and further improve the accuracy and robustness of the results through their complementary attributes.

3.6. Team: dashengge

Members: Runsheng Wang (*Huazhong University of Science and Technology*, wrsh@hust.edu.cn), Runyu Wang (*University of Southern California*, runyuwan@usc.edu), Shijuan Huang (*Huazhong University of Science and Technology*, hshijuan@qq.com), Jianbo Li (*Huazhong University of Science and Technology*, m202273875@hust.edu.cn), Zongyi Li (*Huazhong University of Science and Technology*, D202081087@hust.edu.cn)
Method: In the solution, three models, SwinGait [3], DeepGaitV2 [3], and BNBaseline, are involved. The implementation details are available in OpneGait [4]. The framework of the solution is shown in Figure 5. Three models were trained, and they are BNBaseline, DeepGaitV2, and SwinGait. BNBaseline is the Baseline model offered in the earlier version of the OpenGait project without

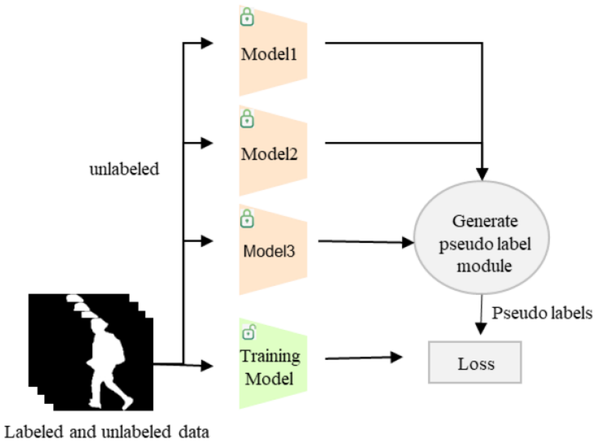

Figure 4. The framework of Team BRAVO-FJ's method.

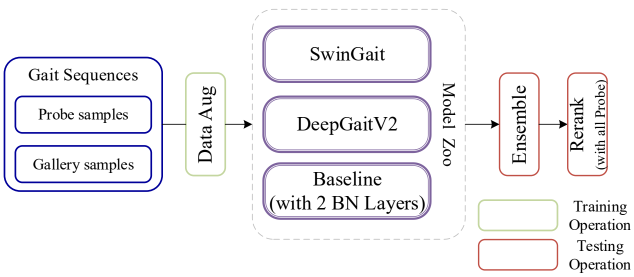

Figure 5. The framework of Team dashengge. Three models, including SwinGait, DeepGaitV2 [3], and BNBaseline, are involved. The green box and red boxes denote the operations used in the training and testing phase, respectively.

the residual learning. Moreover, each of the last two convolution layers is followed by one batch normalization layer. The three models were trained on the HID2022 dataset. During the training stage, three augmentation strategies are used, including the horizontal flip, random rotation and perspective. During the test stage, the three models were ensembled, and all probe data was involve both in phase 1 and phase 2. This strategy can boost the results after re-ranking [20]. The features of and distance matrices of probe and gallery samples are generated with the three models, which are summed as the final distance matrix for ensembling the three models.

Ablation studies are conducted to demonstrate the effectiveness of various models and strategies. The results of these ablation studies are presented in Table 3. After ensembling all three models and involving all probe samples for reranking, the framework achieved the best result.

3.7. Team: HUST-MCLAB

Members: Yunfei Ni, Haijun Xiong, and Bin Feng
Huazhong University of Science and Technology
{U202211792, xionghj, fengbin}@hust.edu.cn

Method: The method is shown in Figure 6, and it comprises three core components: data pretreatment, training, and test.

(1) Data Pretreatment: The data pretreatment utilizes a Maximum Connectivity graph (MC) to remove noise and redundant data from the probe images, as illustrated in Figure 7. This step enhances the model's ability to accurately perceive the probe data. Additionally, data is resized to 64×64 pixels to match the input requirements of the models.

Table 3. Ablation study of different models and strategies. "RR" means re-ranking. "RR all" means re-ranking with the probe samples of phase 1 and phase 2.

DeepGait	SwinGait	Baseline	RR	RR all	Acc
✓			✓		61.8%
✓		✓	✓		64.7%
✓	✓		✓		65.2%
✓	✓	✓	✓		66.8%
✓	✓	✓		✓	68.2%

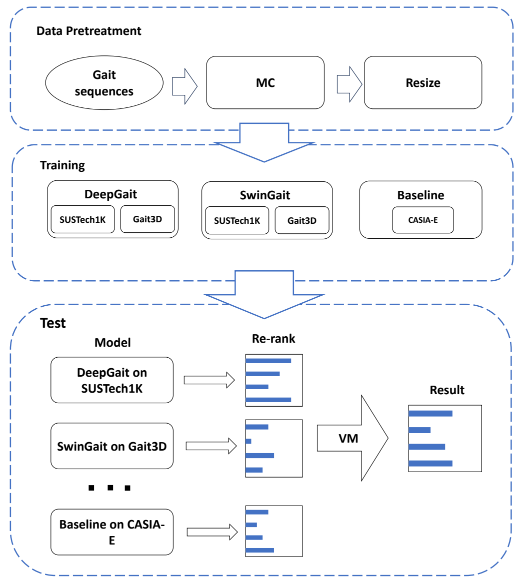

Figure 6. The framework of Team HUST-MCLAB's method.

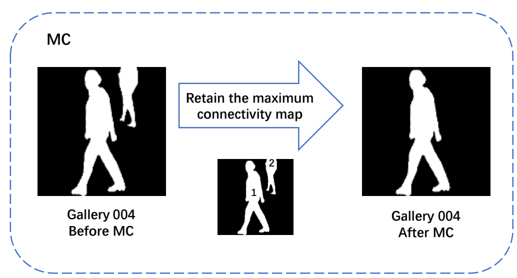

Figure 7. The overview of the MC block.

(2) Training: The training is conducted with OpenGait [4], where various models are trialed across different datasets to identify the most effective. Particularly, the models DeepGait [3] and SwinGait [3], trained on the SUSTech1K [12] and Gait3D [19] datasets, show superior cross-domain capabilities. The reason may be that the two datasets were collected from real-world scenarios. Notably, the DeepGait model trained on Gait3D is adjusted to an input size of 128×128 pixels, contrasting with the 64×64 size used by other models.

(3) Test: During the test phase, a Re-ranking (RK) and Vote Mechanism (VM) are employed to enhance the robustness of the results. RK is utilized during the feature extraction stage of each model to bolster feature recognition. Concurrently, VM integrates the capabilities of multiple models, improving the overall accuracy and robustness of the outputs.

Performance: Results from ablation studies indicate that RK significantly enhances model accuracy. As demonstrated in Table 4, models such as DeepGait and SwinGait trained on SUSTech1K and Gait3D, along with Baseline [4] trained on CASIA-E [13], achieve high accuracies. The integration of VM notably enhances the performance across multiple models, corroborating the effectiveness of this combined approach.

Table 4. The rank-1 accuracies by different models

Model & Dataset	Accuracy (%)
DeepGait on SUSTech1K	60.50
DeepGait on Gait3D	60.86
SwinGait on SUSTech1K	62.08
SwinGait on Gait3D	55.53
Baseline on CASIA-E	55.79
VM	66.45

4. Analysis

The technologies used by the top teams are summarized in Table 1. Similar with HID 2023 [18], the technologies of data cleaning, data alignment, data augmentation, re-ranking and ensemble played important roles in HID 2024.

Even we had a more strict role this year that the test data in HID 2023 and HID 2024 cannot be used in training, a much higher best accuracy 84.9% was achieve in HID 2024 (80.8% in HID 2023). The improvements this year should be brought by the strong backbone models. Most top teams (7 of 8) chose DeepGaitV2 as the backbone model. It means a strong backbone model for gait recognition is essential.

Surely DeepGaitV2 is not the best in the future. To have a better backbone for gait recognition, we may need a much larger dataset to train it. But it is very difficult to collect gait data and label the data. Some large vision models trained using images or videos (not gait data) can be applied for gait recognition as in [8, 15].

5. Conclusions and Future Paths

The 5th International Competition on Human Identification at a Distance (HID 2024) marks a significant milestone in the evolution of gait recognition. Building upon the foundations laid by previous iterations, this year's competition introduced more complex challenges, including the cross-domain *SUSTech-Competition* dataset featuring 859 subjects with various clothing, carrying conditions, occlusions, and view angles. The results demonstrate remarkable progress, with the top-performing team achieving an impressive 84.9% accuracy on this challenging dataset.

The methods employed by leading teams reveal a convergence towards sophisticated deep learning architectures, particularly variants of DeepGaitV2 and SwinGait models [3]. These approaches, combined with innovative data augmentation techniques, ensemble methods, and cross-domain adaptation strategies, highlight the increasing maturity of gait recognition systems. The diverse range of techniques employed, from advanced preprocessing methods to novel network architectures, underscores the multifaceted nature of the challenge and the creativity of the research community.

With 90 registered teams, HID 2024 attracted great interest in this field. The competition's outcomes suggest that gait recognition is approaching a level of robustness and reliability suitable for real-world applications, marking a significant step forward in biometrics. This progress not only advances the state-of-the-art in gait recognition but also paves the way for broader applications in security, healthcare, and smart environments.

In the future, we need to explore multi-modal and context-aware gait recognition along with corresponding datasets. Can we integrate multiple modalities beyond silhouettes, such as 3D point clouds, skeletal data, and contextual information, in challenging real-world scenarios? Investigating the fusion of these diverse data types and developing context-aware models that can adapt to varying environmental conditions will be crucial for advancing the field. However, challenges remain in developing a robust dataset in this arena.

Thus far, we have not incorporated privacy-preserving and ethical gait analysis. We feel that researchers should concentrate on developing privacy-preserving techniques that can operate effectively while ensuring the protection of individuals' identities. Moreover, establishing ethical guidelines and standards for the deployment of gait recognition systems in public spaces will be essential for responsible advancement of the technology.

Future competitions and research efforts should focus on evaluating and improving the robustness of gait recognition systems against adversarial attacks and long-term gait variations. We should also explore the limitations observed in current methods, particularly in enhancing cross-domain capabilities. It requires sincere collaborations between academia and industry to translate these advancements into practical applications, thereby broadening the impact of this research on security, surveillance, and beyond.

Finally, the HID competitions have consistently driven forward the state-of-the-art in gait recognition, fostering both competitive and innovative solutions. As we move forward, the insights and developments from HID 2024 will undoubtedly play a fundamental role in shaping the future of this field. This progress ensures that gait recognition continues to advance in ways that are not only technically impressive but also ethically sound and practically applicable in diverse real-world scenarios.

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 and Mr. Rui Wang. Their assistance and expertise

greatly contributed to the successful organization and execution of the competition.

References

[1] H. Chao, K. Wang, Y. He, J. Zhang, and J. Feng. GaitSet: Cross-view gait recognition through utilizing gait as a deep set. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(7):3467–3478, 2022. 1

[2] P. Connor and A. Ross. Biometric recognition by gait: A survey of modalities and features. *Computer Vision and Image Understanding*, 167:1–27, 2018. 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, 6, 7

[4] C. Fan, J. Liang, C. Shen, S. Hou, Y. Huang, and S. Yu. OpenGait: Revisiting gait recognition toward better practicality. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 1, 3, 4, 5, 6, 7

[5] A. Li, S. Hou, Q. Cai, Y. Fu, and Y. Huang. Gait recognition with drones: A benchmark. *IEEE Transactions on Multimedia*, 26:3530–3540, 2024. 2

[6] W. Li, S. Hou, C. Zhang, C. Cao, X. Liu, Y. Huang, and Y. Zhao. An in-depth exploration of person re-identification and gait recognition in cloth-changing conditions. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 13824–13833, June 2023. 2, 5

[7] B. Lin, S. Zhang, and X. Yu. Gait recognition via effective global-local feature representation and local temporal aggregation. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 14648–14656, 2021. 1, 4

[8] K. Ma, Y. Fu, C. Cao, S. Hou, Y. Huang, and D. Zheng. Learning visual prompt for gait recognition. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 593–603, June 2024. 7

[9] M. S. Nixon, T. Tan, and R. Chellappa. *Human Identification Based on Gait*. Springer, 2006. 1

[10] Z. Qiu, T. Yao, and T. Mei. Learning spatio-temporal representation with pseudo-3d residual networks. In *IEEE International Conference on Computer Vision (ICCV)*, pages 5534–5542, 2017. 5

[11] S. Sarkar, P. Phillips, Z. Liu, I. Vega, P. Grother, and K. Bowyer. The humanid gait challenge problem: data sets, performance, and analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(2):162–177, 2005. 1

[12] C. Shen, C. Fan, W. Wu, R. Wang, G. Q. Huang, and S. Yu. Lidargait: Benchmarking 3d gait recognition with point clouds. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1054–1063, June 2023. 2, 5, 6

[13] C. Song, Y. Huang, W. Wang, and L. Wang. CASIA-E: A large comprehensive dataset for gait recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(3):2801–2815, 2023. 1, 2, 5, 7

[14] 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 Transactions on Computer Vision and Applications*, 10(4):1–14, 2018. 2, 5

[15] D. Ye, C. Fan, J. Ma, X. Liu, and S. Yu. BigGait: Learning gait representation you want by large vision models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 200–210, June 2024. 1, 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 *IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–9, 2022. 5

[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 *International Conference on Pattern Recognition (ICPR)*, volume 4, pages 441–444. IEEE, 2006. 2, 5

[18] S. Yu, C. Wang, Y. Zhao, L. Wang, M. Wang, Q. Li, W. Li, R. Wang, Y. Huang, L. Wang, Y. Makihara, and M. A. R. Ahad. Human identification at a distance: Challenges, methods and results on HID 2023. In *IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–8, 2023. 7

[19] 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 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 1, 2, 5, 6

[20] Z. Zhong, L. Zheng, D. Cao, and S. Li. Re-ranking person re-identification with k-reciprocal encoding. In *IEEE Computer Vision and Pattern Recognition Conference (CVPR)*, pages 3652–3661, 2017. 6

[21] 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 *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021. 1, 2, 5

[22] S. Zou, C. Fan, J. Xiong, C. Shen, S. Yu, and J. Tang. Cross-covariate gait recognition: A benchmark. In *AAAI Conference on Artificial Intelligence (AAAI)*, 2024. 2