Appendix

Anonymous ICCV submission

Paper ID

1. Experiments

1.1. Datasets

CASIA-B. The CASIA-B dataset [9] is the popular cross-cloth gait database. It includes 124 subjects, each of which has 10 groups of videos. Among these groups, six of them are sampled in normal walking (NM), two groups are in walking with a bag (BG), and the rest are in walking in different cloth (CL). Each group contains 11 gait sequences from different angles (0°-180° and the sampling interval is 18°). Therefore, there are 124 (subject) × 10 (groups) × 11 (view angle) = 13,640 gait sequences in CASIA-B. The gait sequence of each subject are divided into training set and test set. Following the setting of previous works [2], in small-sample training (ST) the first 24 subjects (labeled in 001-024) are used for training and the rest 100 subjects are left for testing. In medium-sample training (MT), the first 62 subjects are used for training and the rest 62 subjects are left for testing. In large-sample training (LT), the first 74 subjects are used for training and the rest 50 subjects are left for testing. In the test stage, the sequences NM#01-NM#04 are taken as the gallery set, while the sequences NM#05-NM#06, BG#01-BG#02, and CL#01-CL#02 are considered as the probe set to evaluate the performance.

Outdoor-Gait. The Outdoor-Gait [5] dataset also contains rich clothing and bag variations with complex outdoor backgrounds. Outdoor-Gait contains 138 people with 3 different clothing conditions (NM: normal, CL: different cloth, BG: with bag) in 3 scenes (SCENE-1: simple background, SCENE-2: static and complex background, SCENE-3: dynamic and complex background with moving objects). Following the setting of previous works [1], we take the first 69 subjects for training and the left subjects for testing and for each condition, there are at least 2 video sequences in gallery and probe.

OUMVLP. The OUMVLP dataset [6] is one of the largest view-variation gait dataset. This dataset includes more than 10,000 subjects. Each subject's sequences are captured under 14 views (0°-270° and the sampling interval is 15°). There are two sequences under each view. There is no CL or BG walking condition.

1.2. Construction of Noisy Gait Datasets

Noise type. In order to tackle the degradation caused by various sources of noise in labels and appearances, with the ultimate aim of mitigating label noise memorization and appearance memorization, we establish these distinct settings.

1. **Random Label Noise.** Following the settings outlined in the study [7] on constructing noisy CIFAR datasets, we generate a random label noise dataset by flipping labels to alternative categories across the entire id lists.

2. **Clothing Label Noise.** This particular noise setting aims to emulate the label noise observed in monitoring and detection-based acquisition of data, wherein a single pedestrian is split into two distinct identities. For instance, from monitor A, a suspect is denoted as '001', while from monitor B, the same suspect with different clothes is denoted as another identity. Also, in video clip A, a pedestrian is denoted as '001', while in video clip B, the same individual, albeit wearing different clothing, is identified as, e.g. '007'. To realize this emulation, we disentangle the gait sequences of '001' NM, BG, and CL into ('001' NM, BG) and ('007' CL).

3. **Appearance Noise/Augmentation Noise.** In order to replicate the data noise introduced through preprocessing techniques, as depicted in Figure 3, we generate appearance disturbances by employing morphological operations, specifically dilate and erode operations. The noise rates are set to 0.1 and 0.2, while the default positions for the dilate/erode operations are determined as [8,56] for the gait sequence with a resolution of 64 and [16,112] for the gait sequence with a resolution of 128.

Construction of noisy CASIA-B. We establish two noisy variations of CASIA-B, namely Noisy Clothing CASIA-B and Noisy-CASIA-B.

Noisy Clothing CASIA-B exclusively focuses on Clothing Label Noise, as previously described, where a pedestrian is split into two separate identities. For instance, '001 NM#01' is transformed into '001NM#01', and '001 CL#01'

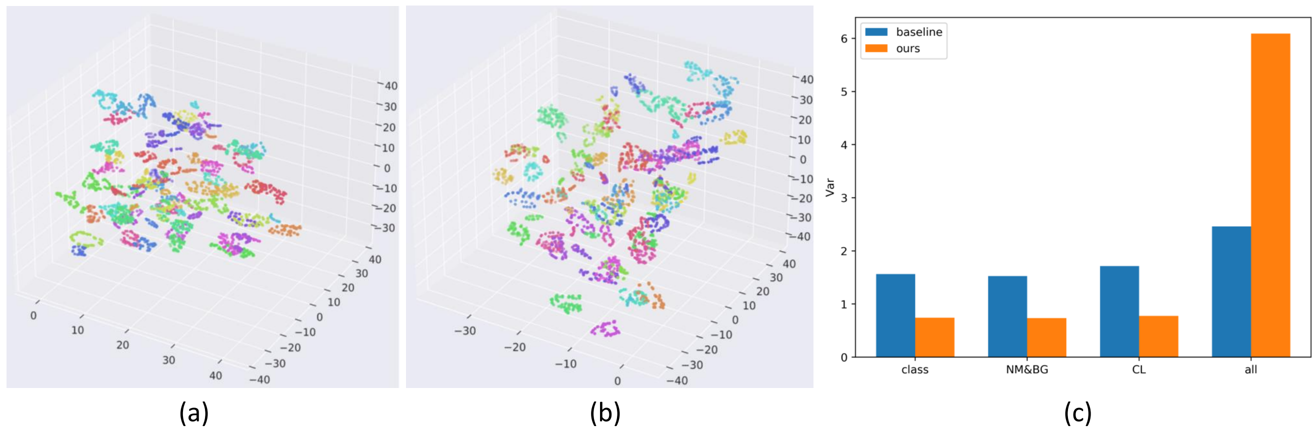

Figure 1. Analysis on test set. (a) The TSNE visualization of baseline features. (b) The TSNE visualization of CNTN features. (c) The average variance of features. ‘class’ denotes the average variance of features from the same class. ‘NM&BG’ denotes the average variance of features from the same class and the walking conditions in NM and BG. ‘CL’ denotes the average variance of features from the same class and the walking conditions in CL. ‘all’ denotes the variance of all test features.

becomes ’130NM#01’. Notably, this particular configuration is generated with a noise percentage of 0.6, wherein the noisy instances are limited to ’001’-’044’ in the training set. Noisy Clothing CASIA-B represents the most practical scenario for gait recognition, as in real-world applications, it is often the varying clothing worn by pedestrians that confounds both human annotators and unsupervised clustering methods, leading to inaccurate labeling.

Noisy-CASIA-B with Random Label Noise, is constructed by making the noise percentage of original gait sequences labeled with random noisy labels in this dataset, regardless of its walking conditions (eg, ’001 NM#01’ → ’002 NM#01’, ’001 CL#01’ → ’002 CL#01’). The noise rate is set to 0.1 and 0.2 in this paper.

Noisy-CASIA-B with Augmentation Noise corresponds to the appearance of random disturbance which simulates the preprocessing noise, containing morphological operations. The noise rate is set to 0.1 and 0.2 in this paper.

Construction of noisy Outdoor-Gait. The noise settings remains the same as construction of noisy CASIA-B.

1.3. Evaluation metrics

In the testing phase, we compare the feature similarities between probe and gallery samples to identify a person and report performance of the average Rank-1 recognition accuracy.

1.4. Noisy Settings

The experiment settings in this paper are twofold, including experiments on three benchmarks and experiments on two reconstructed noisy datasets. Experiments on three benchmarks just follows everything the same in previous works [4, 2], and the results of CASIA-B, Outdoor-Gait and OUMVLP are in Tab.**??**, Tab.**??** and Tab.**??** respectively. Experiments on noisy settings includes normal noise setting and noisy clothing setting. Normal noise setting follows the noise setting in other noisy tasks (noisy image classification on CIFAR10 and CIFAR100) [3, 8], and results are in Tab.**??**. However, this random noise setting is not close to practical situations in gait recognition task since the annotators and the unsupervised clustering methods have a clear tendency that the pedestrians with different clothing are misclassified. Thus, to better simulate the practical clothing noise, we also includes a noisy clothing setting, and the results of Noisy Clothing CASIA-B and Noisy Clothing Outdoor-Gait are in Tab.**??** and Tab.**??** respectively.

1.5. Visualization Feature Representations on Noisy-CASIA-B

Although TSNE is just one way to compact the features to visualized dimensions, it can still provide us with some intuitional feelings. In Fig.1, CNTN learns a larger space than baseline, while at the same time, every class cluster is more compact and aggregated. Fig.1 (c), the variance on the test set can also support this. The average class variance nearly halves the baseline, while the total variance triples the baseline. The larger space and better feature distribution is one of the reflections that the CNTN learns a pattern with less memorization and better generalization.

References

[1] Hanqing Chao, Yiwei He, Junping Zhang, and Jianfeng Feng. Gaitset: Regarding gait as a set for cross-view gait recognition. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 8126–8133, 2019. 1

[2] Hanqing Chao, Kun Wang, Yiwei He, Junping Zhang, and Jianfeng Feng. Gaitset: Cross-view gait recognition through utilizing gait as a deep set. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021. 1, 2

[3] Bo Han, Quanming Yao, Xingrui Yu, Gang Niu, Miao Xu, Weihua Hu, Ivor Tsang, and Masashi Sugiyama. Co-teaching: Robust training of deep neural networks with extremely noisy labels. *Advances in neural information processing systems*, 31, 2018. 2

[4] Beibei Lin, Shunli Zhang, and Feng Bao. Gait recognition with multiple-temporal-scale 3d convolutional neural network. In *Proceedings of the 28th ACM international conference on multimedia*, pages 3054–3062, 2020. 2

[5] Chunfeng Song, Yongzhen Huang, Yan Huang, Ning Jia, and Liang Wang. Gaitnet: An end-to-end network for gait based human identification. *Pattern recognition*, 96:106988, 2019. 1

[6] Noriko Takemura, Yasushi Makihara, Daigo Muramatsu, Tomio Echigo, and Yasushi Yagi. Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition. *IPSJ Transactions on Computer Vision and Applications*, 10(1):1–14, 2018. 1

[7] Hongxin Wei, Lei Feng, Xiangyu Chen, and Bo An. Combating noisy labels by agreement: A joint training method with co-regularization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13726–13735, 2020. 1

[8] Xingrui Yu, Bo Han, Jiangchao Yao, Gang Niu, Ivor Tsang, and Masashi Sugiyama. How does disagreement help generalization against label corruption? In *International Conference on Machine Learning*, pages 7164–7173. PMLR, 2019. 2

[9] Shuai Zheng, Junge Zhang, Kaiqi Huang, Ran He, and Tieniu Tan. Robust view transformation model for gait recognition. In *2011 18th IEEE International Conference on Image Processing*, pages 2073–2076, 2011. 1