Appendix

Anonymous ICCV submission

Paper ID

1. Experiments

1.1. Datasets

CASIA-B. The CASIA-B dataset [8] 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. To address the noise that appears in labels and appearances and thus alleviate the corresponding label noise memorization and appearance memorization, respectively, we construct two noise settings including noisy labels and noisy data. **Random Label Noise.** As the noisy CIFAR, we construct the random label noise dataset by randomly flip the label to all other labels in the dataset. **Clothing Label Noise.** This noise setting is to mimic the noise in the monitor and detection, where one pedestrian is separated to two identities and creates a new identity. For example, in video clip A, a pedestrian is labeled as '001', and in video clip B, the same pedestrian with differnt clothing is labeled as '100'. To achieve this, we separate the gait sequences of '001'NM, BG, CL to '001'NM BG, and '100'CL. **Appearance Noise/Augmentation Noise.** To mimic the data noise caused by preprocessing, as in Fig.3, we construct the appearance disturbance by using morphological operations, e.g. dilate and erode. The noise rate is 0.1 and 0.2, and the default dilate/erode position is [8,56] for images with the resolution of 64, and [16,112] for images with the resolution of 128.

Noisy-CASIA-B. We construct Noisy-CASIA-B into two different settings. The first is the random noise setting with noise rate 0.1 and 0.2 which contains random noise and augmentation noise. Random noise corresponds to the random label noise, which indicates that the noise percentage of original gait sequences are 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'). Appearance noise corresponds to the appearance random disturbance which simulates the clothing change, containing morphological operations. The second is noisy clothing setting where one pedestrian is separated to two identities and creates a new identity (eg, '001 NM#01' → '001NM#01', '001 CL#01' → '130NM#01'), with a percentage of 0.6 (In train set, the whole dataset is '001'-'074', noisy ones are '001'-'044'). Noisy Clothing CASIA-B is the most practical setting for gait recognition since in industry, it is usually the pedestrians with different clothing that makes both annotators and unsupervised clustering methods confused and thus

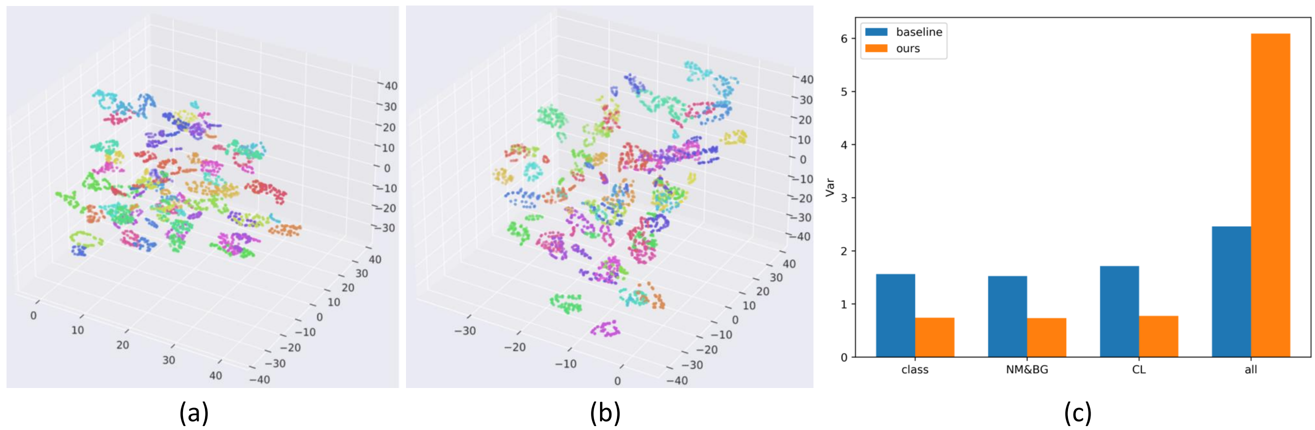

Figure 1. Analysis on test set. (a) The TSNE visualization of baseline features. (b) The TSNE visualization of CNTN features. (c) The average variance of features. ‘class’ denotes the average variance of features from the same class. ‘NM&BG’ denotes the average variance of features from the same class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

with incorrect labels.

Noisy-Outdoor-Gait. We construct Noisy-Outdoor-Gait which accords with 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, 7], 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] 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

[8] 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