Supplementary Material

Joint Generative and Contrastive Learning for Unsupervised Person Re-identification

Appendices

Appendix A. Cycle consistency

Two kinds of cycle consistency are used in our proposed method GCL. 1) Based on a popular assumption in ReID that two images of a same identity should have same nearest neighbors in the dataset, we have calculated k-reciprocal Jaccard distance [4] for the DBSCAN clustering. This operation effectively makes pseudo labels more reliable for contrastive learning. 2) Since no paired data are available, we have used the CycleGAN [5] structure to supervise the generative module. By minimizing the image and feature reconstruction losses in a cycle consistency, representations are disentangled into appearance and structure features, which permits generating person images in novel view-points without changing identity information.

Appendix B. View-invariant losses

We illustrate another example in Fig. 1 to confirm the effectiveness of the view-invariant losses in generation. When GCL is trained without the view-invariant losses, GCL degrades to a traditional CycleGAN, which is prone to affected by the noise inside the original image. The view-invariant losses help the identity encoder to extract identity features shared between different views, which are robust to the noise inside the original image.

Appendix C. Effects on pseudo labels

We minimize intra-class variance via contrasting generated images, which leads to a larger inter-class distance in latent space. Learning view-invariant representations from diversified generated data helps clustering algorithms to generate more accurate pseudo labels. With a same DBSCAN clustering, the cluster number of GCL is closer to real identity number than that of contrastive learning with traditional data augmentation. For example, Market-1501 dataset has 751 real identities. DBSCAN in GCL categorizes unlabeled images into around 520 clusters, while the contrastive learning with traditional data augmentation has around 460 clusters (see Fig. 2).

Figure 1. More qualitative ablation study on the view-invariant losses. For simplicity, $\mathcal{L}_{vi}$ denotes three view-invariant losses $\mathcal{L}_{vi}+\mathcal{L}'_{vi}+\mathcal{L}''_{vi}$, which helps E_{id} to extract better identity features (white shirt).

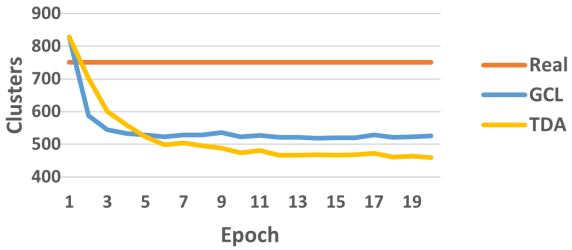

Figure 2. Cluster number curve on Market-1501. TDA denotes traditional data augmentation, including random flipping, cropping, jittering, erasing.

Appendix D. Generated views

We illustrate more examples of generated views with a JVTC [3] fully unsupervised baseline on Market-1501 in Fig. 3, DukeMTCM-reID in Fig. 4 and MSMT17 in Fig. 5. Here, we show generated examples from both training and test sets to confirm the effectiveness of our GCL. Gener-

	45°	90°	135°	180°	225°	270°	315°
Market	56.55	75.83	62.59	51.22	62.31	70.79	55.00
Duke	58.24	72.21	66.29	57.41	64.61	68.08	55.03
MSMT	54.14	64.46	60.75	55.98	59.78	62.26	48.40

Table 1. FID score on different views.

ally, the generation quality is good enough to help our GCL learn view-invariant representations. However, there are still some limitations, *e.g.*, some visual blurs still exist and detailed identity information is lost in some cases (in the *bottom row* of Fig. 3, the red logo on the shorts disappears in the generated images). In future work, we believe that the visual blurs can be alleviated by leveraging the architectures from more recent GANs [1, 2] in our generator and detailed identity information can be better preserved when better unsupervised baselines are available.

Generally, it is easier to generate novel views of $45°$, $180°$ and $315°$. $45°$ and $315°$ are small rotations, in which original and synthesized images can share maximal identity information. $180°$ can be roughly regarded as a horizontal flipping. Results in Tab. 1 verify this supposition. Our generated novel views on the three datasets will be released as a new dataset to facilitate future research on view-invariant and unsupervised ReID.

References

[1] Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale GAN training for high fidelity natural image synthesis. In *ICLR*, 2019.

[2] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of StyleGAN. In *CVPR*, 2020.

[3] Jianing Li and Shiliang Zhang. Joint visual and temporal consistency for unsupervised domain adaptive person re-identification. In *ECCV*, 2020.

[4] Zhun Zhong, Liang Zheng, Donglin Cao, and Shaozi Li. Re-ranking person re-identification with k-reciprocal encoding. In *CVPR*, 2017.

[5] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *ICCV*, 2017.

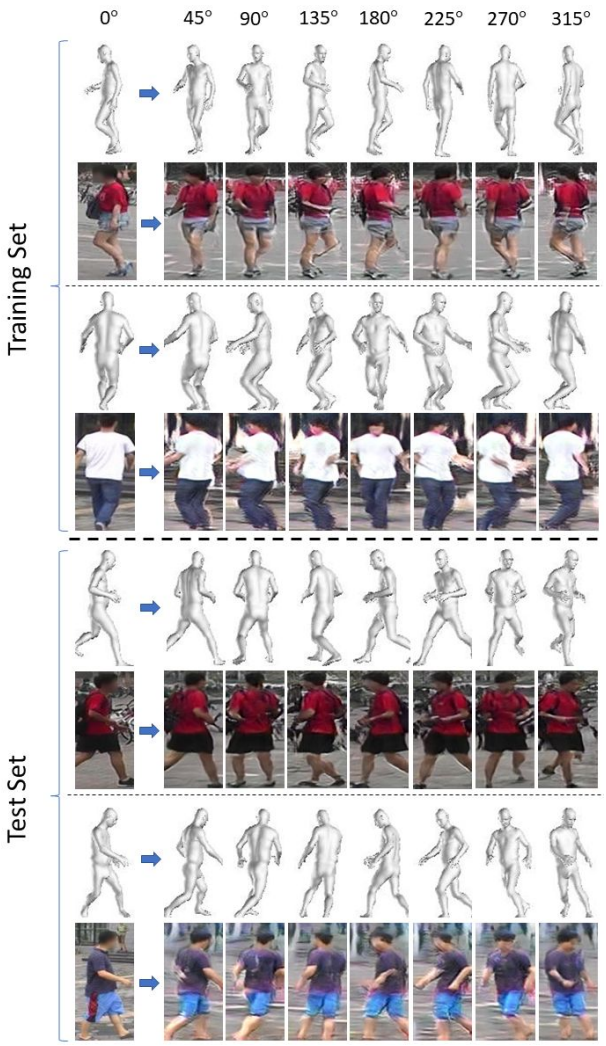

Figure 3. Examples of generated novel views on **Market-1501** training and test sets.

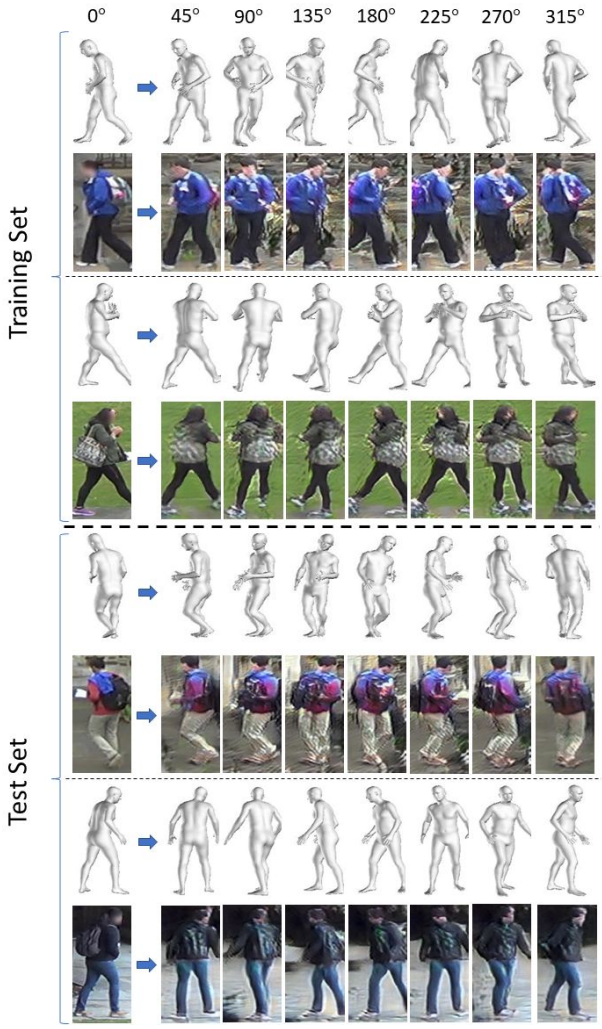

Figure 4. Examples of generated novel views on **DukeMTMC-reID** training and test sets.

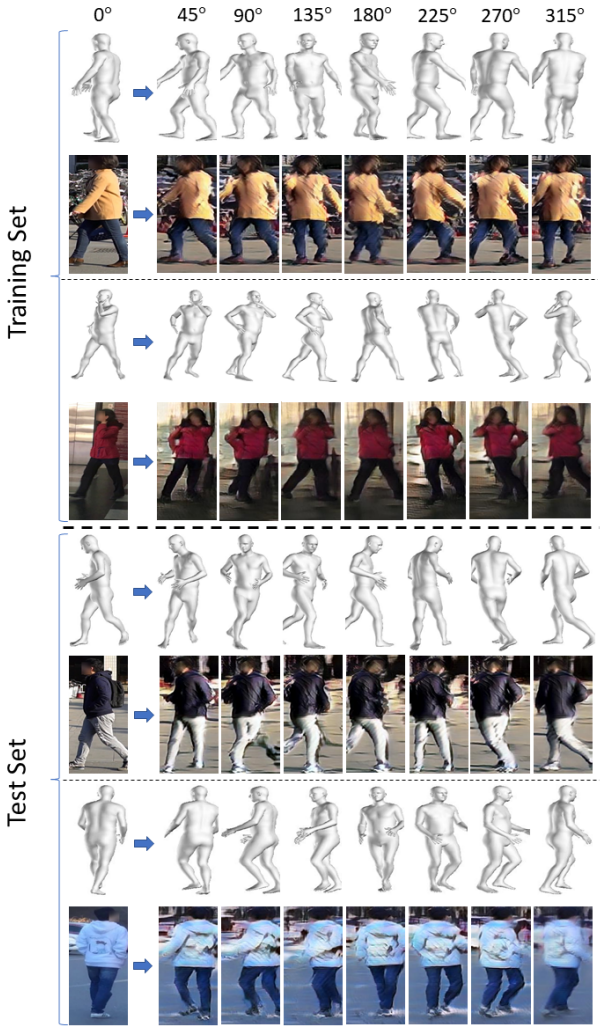

Figure 5. Examples of generated novel views on **MSMT17** training and test sets.