IFS/ACM Transactions on Machine Learning VOL. 1, NO. 1 (OCTOBER 2024) PP. 7-12

Fusion Cycle GAN: A Multiple Feature Fusion Based Cycle-consistent Generative Adversarial Network for Person Re-identification

Hang Li
Software Collage
Shenyang Normal University
Shenyang, Liaoning, China

Chu Zhao
Software Collage
Shenyang Normal University
Shenyang, Liaoning, China

Editor: Khuhro Mansoor

Abstract

In the domain of person re-identification, conventional approaches are often susceptible to interference from shifts of variables such as background, veils, and attire, leading to a degradation in identification accuracy. To mitigate the influences of these shifts on identification performance, this paper introduces a novel person re-identification approach that leverages cycle-consistent generative adversarial networks coupled with multiple feature fusion. This method achieves re-identification by assessing and comparing the measured distances between pairs of persons. The approach is bifurcated into two streams: one for extracting global features and another for capturing local features. Subsequently, these global and local features are integrated. The integrated features are then subjected to contrast metric distance learning, where similarity scores are computed to rank the samples. Extensive experimental outcomes on substantial datasets such as CUHK03 and VIPER demonstrate that this innovative method effectively diminishes the influence of background, veils, attire, and other alterations on identification accuracy.

Keywords: person re-identification, multiple feature fusion, cycle-consistent generative adversarial network

1 Introduction

Person re-identification is the process of confirming whether individuals captured by two separate cameras are the same person within the non-overlapping surveillance areas of the cameras, as shown in Figure 1. This technology holds immense potential for application in criminal investigations, residential security management, and life scenarios like searching for missing elderly or children.

Over the recent past, significant strides have been taken in the field of computer vision, particularly in interpreting behavior through the use of visual sensor networks. The area of person re-identification has garnered increasing interest from the research community. The majority of conventional research approaches have focused on accomplishing person

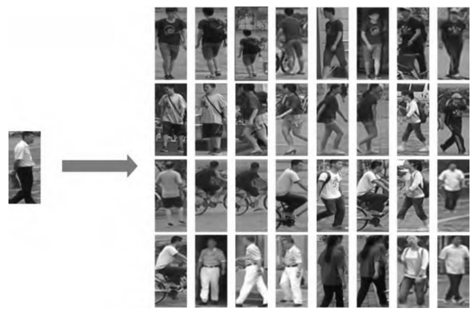

Figure 1: Person re-identification

re-identification by leveraging low-level features, including color, texture, and attire. For example, pioneering studies on color-based recognition methods extracted color features from pixel data across two cameras to facilitate the re-identification of persons.

Consequently, the identification outcomes are distinguished along a single dimension. This implies that the identification of distinct persons relies on a single characteristic. There are studies that amalgamate an array of features for the re-identification of persons. Yet, their effectiveness is constrained by the computational complexity that arises from the creation of high-dimensional feature spaces.

Owing to a multitude of challenges such as limited frame resolution, variations in lighting and posture, occlusions, camera perspectives, and similarities in appearance, traditional approaches struggle to attain optimal performance. To enhance the outcomes in person re-identification, scholars have begun to explore the application of metric learning. Currently, a variety of metric learning-based person re-identification methods have been investigated, such as KISSME (Tao et al. (2016)) and RLML (Liong et al. (2015)). However, those methods focused on the extraction of local features, and ignored the influences of extracting global features.

To address the aforementioned problems, this study introduces a person re-identification approach that leverages cycle-consistent generative adversarial networks coupled with the fusion of multiple features. The methodology is divided into two main streams: global feature extraction and local feature extraction. Once the global and local features have been extracted, they are integrated into a cohesive representation for contrast metric distance learning and ranking.

2 Method

2.1 Global Feature Extraction

By establishing generators and discriminators, generative adversarial networks (GANs) facilitate the mutual learning and competitive interaction between two distinct image styles, thereby continuously enhancing their respective capabilities for generation and discrimination. Ultimately, this leads to the transformation of one image style into another with highly realistic results. While GAN networks have numerous variations, the Cycle GAN network utilizes convolutional and deconvolutional strategies, and it does not necessitate paired image inputs, which contributes to the superior quality of the generated images (Wu

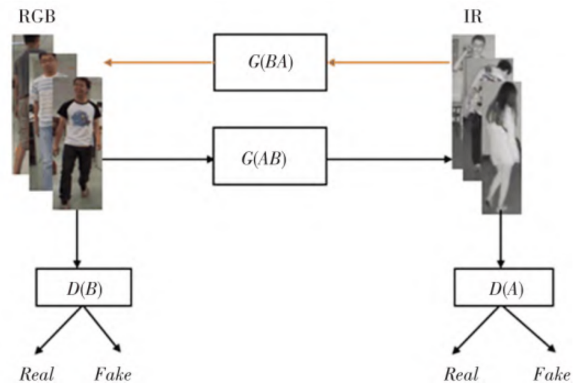

Figure 2: The Cycle GAN network to facilitate the transformation and augmentation between RGB and IR image formats

et al. (2023); Li et al. (2023)). Consequently, this study employs the Cycle GAN network to extract the global features via facilitating the transformation and augmentation between RGB and IR image formats, as shown in Figure 2.

The Cycle GAN network facilitates the transformation and augmentation of RGB images into IR images, and vice versa, through the following process. The RGB images are fed into the generator $G(AB)$ to produce IR images, while the original IR images are subjected to the discriminator $D(A)$. The authentic IR image is identified as "Real", whereas the generated IR image is labeled as "Fake". Subsequently, the "Fake-IR" image is processed by the generator $G(BA)$ once more to create Cycle-RGB images, which should closely match the original RGB images. In a similar fashion, IR images are converted into "Fake-RGB" images and Cycle-IR images. Both the "Fake-RGB" and "Fake-IR" images are intended to be nearly indistinguishable from their "Real" counterparts.

The loss function of cycle GAN includes two sub-losses, namely, GAN loss and cycle-consistent loss. GAN losses of generator G and discriminator D are shown in Equation 1 and Equation 2, respectively, where $||\cdot||_1$ is the L1-form. $Tdata$ and $Fdata$ are true data and fake data, respectively.

$$L_{GAN-G} = E_{Tdata}[|| \log(D_{Tdata})||_1] + E_{Fdata}[|| \log(1 - D_{Fdata})||_1]. \tag{1}$$

$$L_{GAN-D} = E_{Fdata}[|| \log(D_{Fdata})||_1] + E_{Tdata}[|| \log(1 - D_{Tdata})||_1]. \tag{2}$$

Cycle-consistent loss is shown in Equation 3.

$$L_{cyc} = E_{Tdata}[||Tdata - cyc(Tdata)||_1] \tag{3}$$

where $cyc()$ is the combination of $G(AB)$ and $G(BA)$.

2.2 Local Feature Extraction

The region of interest pertains to the distinct local features of persons that markedly set them apart from others, such as distinctive backpacks or notable clothing colors. Presently,

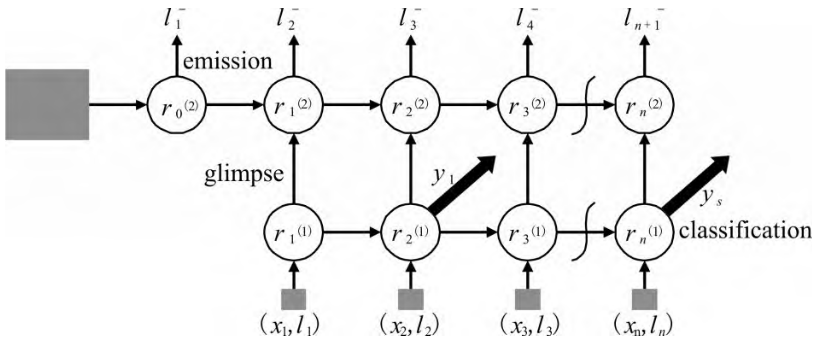

Figure 3: Deep recurrent attention model

models based on LSTM neural networks and RNN neural networks are prevalent. In this study, a deep recurrent attention model is selected to identify and extract the local features in the region of interest of persons, as shown in Figure 3.

2.3 Multiple Feature Fusion

The aim of multiple feature fusion is to balance the effect of local and global features. In this study, to remain the information in local and global features as much as possible, the concatenation fusion is chosen.

2.4 Contrast Metric Distance Learning

After the multiple feature fusion, contrast metric distance learning is conducted via first computing the Mahalanobis distance:

$$d_M = (x_1^R - x_2^R)^T M (x_1^R - x_2^R). \tag{4}$$

where M is a uniform metric. x_1^R and x_2^R are the fusion features of a pair of persons.

Then, the similarity score can be computed:

$$S_{(x_1^R, x_2^R)} = \frac{x_1^R, x_2^R}{|x_1^R| \cdot |x_2^R|} \cdot \frac{1}{d_M}. \tag{5}$$

Finally, all similarity scores are sorted, and the highest one indicates the same person.

3 Experiment

3.1 Experimental Setting

VIPER (Müller et al. (2016)) and CUHK03 (Zhang et al. (2018)) are used as the experimental datasets. The former includes 632 pairs of person images under different lighting conditions. The latter includes 13,164 images of 1360 persons from two cameras in a university campus. Accuracy is utilized as the evaluation metric of experimental results. The comparison methods includes CMCL (Wen et al. (2023)), LCNN (Ke et al. (2021)), and AMR (Li et al. (2021)).

Table 1: Results on VIPER dataset (Accuracy/%)

Method	Rank1	Rank5	Rank10	Rank20
CMCL	44.3	74.2	86.0	94.1
LCNN	40.7	72.4	85.4	91.4
AMR	34.4	70.3	77.4	91.2
Proposed	45.2	75.1	90.3	96.3

Table 2: Results on CUHK03 dataset (Accuracy/%)

Method	Rank1	Rank5	Rank10	Rank20
CMCL	51.4	81.3	88.5	95.9
LCNN	49.9	75.8	88.4	93.1
AMR	41.4	72.5	85.1	93.5
Proposed	52.2	80.5	89.7	96.1

3.2 Experimental Results

According to Table 1 and Table 2, the proposed person re-identification approach outperforms the three comparison methods on the two experimental datasets.

4 Conclusion

Person re-identification presents a complex challenge, often influenced by fluctuations in background, attire, and so on. To mitigate the influences of such changes on identification accuracy, this study introduces a novel person re-identification approach that leverages cycle-consistent generative adversarial networks coupled with multiple feature fusion. The proposed method encompasses two principal streams: local feature extraction and global feature extraction. By employing a deep recurrent attention model, this paper identifies the critical region of interest on a person and subsequently extracts and amalgamates local features from that region with global features to create a comprehensive fusion feature for contrast metric distance learning and ranking. This approach enhances the significance of distinctive local features and diminishes the influences of background, attire, and so on. Experimental results demonstrate that the proposed method has achieved promising outcomes on both datasets.

Acknowledgments and Disclosure of Funding

The authors declare that there is no conflict of interest and no funding.

References

Xiao Ke, Xinru Lin, and Liyun Qin. Lightweight convolutional neural network-based pedestrian detection and re-identification in multiple scenarios. *Machine Vision and Applications*, 32(2):46, 2021.

Chao Li, Xiaoyu Yang, Kangning Yin, Yifan Chang, Zhiguo Wang, and Guangqiang Yin. Pedestrian re-identification based on attribute mining and reasoning. *IET Image Processing*, 15(11):2399–2411, 2021.

Peng Li, Asif Ali Laghari, Mamoon Rashid, Jing Gao, Thippa Reddy Gadekallu, Abdul Rehman Javed, and Shoulin Yin. A deep multimodal adversarial cycle-consistent network for smart enterprise system. *IEEE Transactions on Industrial Informatics*, 19 (1):693–702, 2023.

Venice Erin Liong, Jiwen Lu, and Yongxin Ge. Regularized local metric learning for person re-identification. *Pattern Recognition Letters*, 68:288–296, 2015.

Peter Müller, M Schwerhoff, and A J Summers. Viper: A verification infrastructure for permission-based reasoning. In *Proceedings of the 17th International Conference on Verification, Model Checking, and Abstract Interpretation*, pages 41–62, St. Petersburg, FL, USA, 2016.

Dapeng Tao, Yanan Guo, Mingli Song, Yaotang Li, Zhengtao Yu, and Yuan Yan Tang. Person re-identification by dual-regularized KISS metric learning. *IEEE Transactions on Image Processing*, 25(6):2726–2738, 2016.

Xiongjun Wen, Xin Feng, Ping Li, and Wenfang Chen. Cross-modality collaborative learning identified pedestrian. *The Visual Computer*, 39(9):4117–4132, 2023.

Shixiang Wu, Chao Dong, and Yu Qiao. Blind image restoration based on cycle-consistent network. *IEEE Transactions on Multimedia*, 25:1111–1124, 2023.

Shizhou Zhang, Qi Zhang, Xing Wei, Yanning Zhang, and Yong Xia. Person re-identification with triplet focal loss. *IEEE Access*, 6:78092–78099, 2018.