Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 7557361, 8 pages
https://doi.org/10.1155/2021/7557361

Research Article

SAN-GAL: Spatial Attention Network Guided by Attribute Label for Person Re-identification

Shaoqi Hou,[1] **Chunhui Liu**,[1] **Kangning Yin**,[2] **Yiyin Ding**,[2] **Zhiguo Wang**,[2] **and Guangqiang Yin**[2]

[1]*School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu, China*
[2]*School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu, China*

Correspondence should be addressed to Guangqiang Yin; yingq@uestc.edu.cn

Received 21 May 2021; Accepted 13 August 2021; Published 30 August 2021

Academic Editor: Yan Huang

Person Re-identification (Re-ID) is aimed at solving the matching problem of the same pedestrian at a different time and in different places. Due to the cross-device condition, the appearance of different pedestrians may have a high degree of similarity; at this time, using the global features of pedestrians to match often cannot achieve good results. In order to solve these problems, we designed a Spatial Attention Network Guided by Attribute Label (SAN-GAL), which is a dual-trace network containing both attribute classification and Re-ID. Different from the previous approach of simply adding a branch of attribute binary classification network, our SAN-GAL is mainly divided into two connecting steps. First, with attribute labels as guidance, we generate Attribute Attention Heat map (AAH) through Grad-CAM algorithm to accurately locate fine-grained attribute areas of pedestrians. Then, the Attribute Spatial Attention Module (ASAM) is constructed according to the AHH which is taken as the prior knowledge and introduced into the Re-ID network to assist in the discrimination of the Re-ID task. In particular, our SAN-GAL network can integrate the local attribute information and global ID information of pedestrians without introducing additional attribute region annotation, which has good flexibility and adaptability. The test results on Market1501 and DukeMTMC-reID show that our SAN-GAL can achieve good results and can achieve 85.8% Rank-1 accuracy on DukeMTMC-reID dataset, which is obviously competitive compared with most Re-ID algorithms.

1. Introduction

The biggest feature of smart city is to make full use of the new generation information technology of all walks of life in the city, so as to improve the efficiency of urban management and the quality of citizens' life. As the representative of the new generation of information technology, Internet of Things technology [1–3] and artificial intelligence technology have been more and more widely used.

As a hot field in artificial intelligence (more specifically, in the field of computer vision.), person Re-ID makes up for the deficiency of face recognition technology in cross-camera surveillance images and has a wide application prospect in intelligent video surveillance fields such as airports and supermarkets. However, due to the differences between different cameras and the characteristics of both rigid and flexible pedestrians, its appearance is easily affected by clothing, scale, occlusion, posture, and perspective, which makes person Re-ID become a hot topic with both research value and challenges in the field of computer vision.

In order to solve the above problems, scholars at home and abroad have made many explorations over these years. The traditional Re-ID algorithm relies on some manual features such as color and texture and measures the correlation by calculating the feature distance [4–6]. Due to the complexity of calculation and poor representational ability, these algorithms based on manual features are gradually phased out. With the development of convolutional neural network (CNN), since 2014, scholars began to use deep learning models to solve the problem of person Re-ID [7, 8].

At present, person Re-ID algorithms based on deep learning are mainly divided into two categories: metric learning and representation learning. Metric learning restricts feature space by designing a distance measurement function, so that

intraclass spacing of pedestrian features is decreased and interclass features are increased. Classical methods such as triplet loss [9], quadruple loss [10], and group consistent similarity learning [11], the key of such methods lies in sample selection, especially the mining of difficult samples.

Different from metric learning, representational learning takes person Re-ID as a classification task and focuses on designing robust and reliable pedestrian feature representation. At present, scholars generally adopt the method of obtaining global features to solve the Re-ID problem; that is, only the pedestrian ID label is used, and the loss function constraint is adopted to make the network automatically learn the features that are more discriminative for different pedestrian IDs from the entire pedestrian images [12]. In order to enhance the adaptability of the model under the scenes of scale, occlusion, and blur, some scholars [13–15] introduced the attention mechanism into the Re-ID task, so as to improve the models' attention to the salient information in the global features of pedestrians, while suppressing irrelevant noises. However, since different pedestrians may have a similar appearance and the same pedestrian varies greatly in different environments, they cannot be correctly matched from the perspective of global appearance alone. Studies show that [16], as a kind of prior knowledge, the attributes of pedestrians (such as gender, whether they wear hats, whether they carry backpacks, etc.) contain rich semantic information and can provide key discriminant information for Re-ID. However, the relevant datasets are not easy to collect because of involving privacy issues [17, 18]. In addition to the pedestrian attribute labels marked by Lin et al. [16] on DukeMTMC-reID [19] and Market1501 [20] on the person Re-ID datasets, the current datasets do not mark the related areas of pedestrian attributes.

In order to solve these problems, we proposed a SAN-GAL network, which combines pedestrian attribute labels and attention mechanism, and can introduce fine-grained attribute features into the Re-ID network for auxiliary discrimination without additional attribute region labeling. The main contributions are summarized as follows:

(1) *Locate the Attribute Area*. in the pedestrian attribute classification network, the attribute labels are used to guide, and the Grad-CAM algorithm [21] is combined to generate AAH

(2) *Obtainment of Attribute Spatial Attention*. in the person Re-ID network, feature maps of different locations and sizes are selected and combined with the corresponding size of attention heat maps generated by the attribute classification network; ASAM is constructed to assist the discrimination of Re-ID task

(3) *Design Dual-Trace Network*. the pedestrian attribute classification network and Re-ID network are trained jointly to achieve the purpose of information interaction and mutual optimization

2. Related Works

2.1. Attribute Recognition for Re-ID. Person Re-ID based on attribute classification can accurately and quickly mark the target pedestrians in the pedestrian database according to the predicted attribute labels. In 2017, Lin et al. [16] proposed an Attribute Person Recognition (APR) joint recognition network in order to improve the overall accuracy of person Re-ID network. This network included an identity recognition convolutional neural network and an attribute classification model, which can predict attributes through identity recognition and at the same time integrate attribute learning to improve the Re-ID network. In particular, Lin et al. also marked pedestrian attribute labels on DukeMTMC-reID and Market1501, which helped domestic and foreign scholars to improve Re-ID performance by pedestrian attributes.

2.2. Attention Mechanism for Re-ID. The essence of the attention mechanism is to imitate the human visual signal processing mechanism, in order to selectively observe the area of interest, while ignoring other noise information. Inspired by this, in the field of image scene, Liu et al. [22] proposed a classic network model HPNet (HydraPlus-Net) with advantages in fine-grained feature recognition based on attention neural network in 2017. It is mainly aimed at enhancing recognition by the feedback of multilayer attention to different layers in multiple directions. In the field of video scene, Li et al. [15] innovatively used multiple spatial attention module and diversified regular terms to ensure that each spatial attention module learned different parts of the body. Based on that, image features in the sequence were fused through the temporal attention module, and problems such as pedestrian occlusion and misalignment in the video sequence were well solved.

3. Method

Firstly, we introduced the overall architecture and logical relationship of the proposed SAN-GAL network in Section 3.1; then, we introduced the generative process of AAH in Section 3.2; finally, we described the construction method of ASAM in Section 3.3.

3.1. Spatial Attention Network Guided by Attribute Label (SAN-GAL). In order to introduce the local features of pedestrian salience into the Re-ID task without adding additional regional annotation, we design SAN-GAL, as shown in Figure 1. As a dual-trace network, SAN-GAL consists of two branches: pedestrian attribute classification network and Re-ID network. Both branches are based on the pretrained ResNet50 [23], in which the attribute classification task provides attribute prior information to assist the discrimination of the Re-ID task.

The general process is as follows:

Firstly, the attribute classification network extracts features in the gradient forward propagation process, and the extracted features are aggregated (BN-FC-Softmax) to connect attribute classification losses.

Then, the activation map output from Softmax layer (in the attribute pretraining module) calculates the AAH on the activation map of different positions and sizes in the

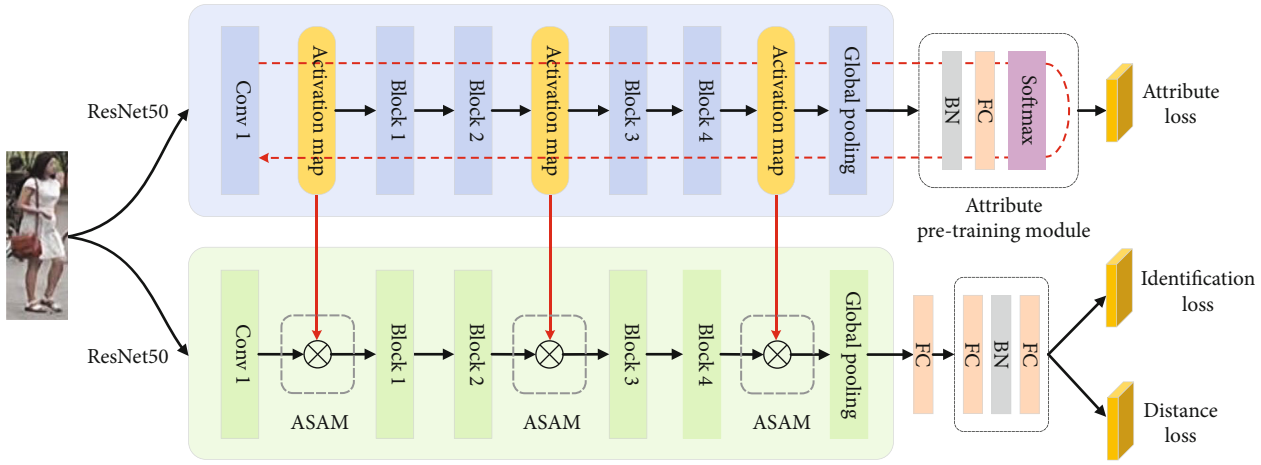

Figure 1: SAN-GAL overall structure diagram.

backbone network through the Grad-CAM algorithm, so as to locate the key area of the attribute.

Finally, in order to enhance the ability of the Re-ID network to extract salient attribute information, the generated AAH is combined with the activation map of the corresponding position in the Re-ID network to construct the ASAM. At the same time, the Re-ID network is optimized after further training.

In particular, in the actual training of SAN-GAL, we have added some special skills:

(1) In the attribute classification network, attribute pre-training module is added only in the first 10 epochs of training and removed after 10 epochs

(2) The two network branches have the same backbone structure but do not share parameters. Since different levels of the network have great differences in the amount of spatial information and semantic information, we introduce the attention mechanism on all three residual blocks with feature map scale changes to enhance the feature processing ability of the model at different fine granularity

(3) In order to achieve the training of the dual-trace network, the attribute classification network needs to be backpropagated twice. After the first calculation of Grad-CAM, the gradient in the activation maps is discarded, and the optimization is achieved after the second update of network parameters. In the actual experiment, in order to improve the computational efficiency, the attribute classification network will complete the pretraining in advance, and the attention parameters calculated offline (from AAH) will be used in the Re-ID network training. In addition to simple implementation, off-line training can also avoid the impact of meaningless AAH at the initial stage of training on the convergence of the Re-ID network, so that the AAH obtained from the same sample calculation is stable and reproducible

3.2. Attribute Attention Heat Map (AAH). Attribute classification network relies on Grad-CAM algorithm to provide spatial focus area for Re-ID. A major advantage of Grad-CAM is that it does not need to transform the model and add additional data annotation, so it is suitable for attention generation algorithm in Re-ID task.

Before calculating Grad-CAM, the output probability y^k predicted by an attribute on Softmax layer in the attribute classification network should be first calculated, and then, the partial derivatives of all pixels on three feature maps of different sizes (i.e., activation maps) on the trunk should be calculated, which can be represented as

$$\frac{\partial y^k}{\partial A_{ij}^c}, \tag{1}$$

where k represents an attribute, (i, j) represents the element coordinates on the current feature map, and c represents the channel of the current feature map.

This result can measure the relevance of some parts of the current feature map to the attribute classification results. Next, the above results are weighted and summed as the coefficients of each channel in the current feature map after global average pooling. After activated by ReLU function, the AAH on the classification of an attribute is obtained, which can be represented as

$$L_{\text{Grad-CAM}}^k = \text{ReLU}\left(\sum_c \left[A^c \times \frac{1}{hw}\left(\sum_{j=1}^{w}\sum_{i=1}^{h}\frac{\partial y^k}{\partial A_{ij}^c}\right)\right]\right), \tag{2}$$

where $L_{\text{Grad-CAM}}^k$ represents the two-dimensional regional heat map generated by the classification attribute k and w and h represent the width and height of the current feature map, respectively. Different pedestrian attributes also have different areas of concern in the current feature map. This concept is shown in Figure 2.

3.3. Attribute Spatial Attention Module (ASAM). The ASAM introduces the attribute prior knowledge into the Re-ID network to assist the discrimination. In order to reduce network complexity and computation consumption, ASAM keeps the structure and size of the activation maps in the Re-ID

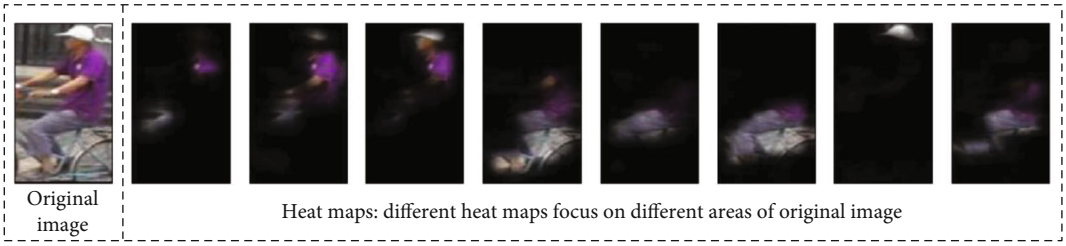

Figure 2: Schematic diagram of heat map generated by Grad-CAM.

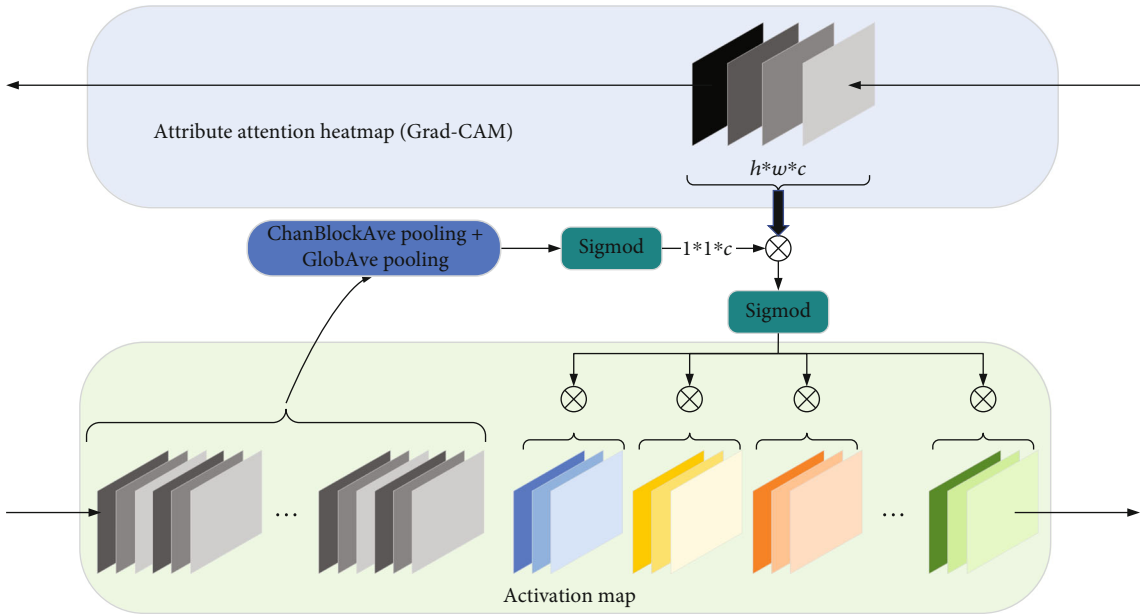

Figure 3: ASAM structure diagram.

network unchanged, as shown in Figure 3. In particular, in each ASAM, the first half of the channel features of the activation map that keeps the corresponding location of the Re-ID network is kept unchanged in order to maintain a certain amount of global information and avoid information loss that may be caused by the attention mechanism. In addition, the first half of the channel features in the activation map was used to learn channel attention, because the contributions of different attributes to the Re-ID task are generally different.

Firstly, the attention parameter β_c of each channel of AHH is calculated, as shown in Equation (3). In order to avoid additional parameters, we use channel block average pooling and global average pooling when we calculate β_c.

$$\beta_c = \text{Sigmoid}\left(\frac{1}{hw}\sum_{k=1}^{c_{\text{attr}}/c_{\text{activ}}}\sum_{j=1}^{w}\sum_{i=1}^{h}x_{ijk}^{c}\right), \tag{3}$$

where (i, j) represents the element coordinate on the current feature map; x_{ijk}^{c} represents the pixel value on this coordinate; w and h, respectively, represent the width and height of the current feature map; $c_{\text{attr}}/c_{\text{activ}}$ represents the number of channels of each block feature map; c_{activ} represents half of the number of channels of the activation map in the corresponding position of the Re-ID network; and c_{attr} represents the number of channels for AAH.

Table 1: Comparison of ablation performance.

Methods	Market1501		DukeMTMC-reID	
	Rank-1 (%)	mAP (%)	Rank-1 (%)	mAP (%)
IDonly (baseline)	92.6	81.3	82.3	72.2
ID+attr	92.4	82.1	82.4	72.4
ID+attrAttention (our SAN-GAL)	94.4	83.9	85.8	74.1

Then, the attention parameters $(1 * 1 * c)$ of all channels are multiplied by each channel of AHH. After activation by Sigmoid function, weighted spatial attention parameter α_c $(h * w * c)$ is obtained, which can be represented as

$$\alpha_c = \text{Sigmoid}\left(\beta_c \otimes L_{\text{Grad-CAM}}^{k}\right). \tag{4}$$

Finally, after the spatial attention parameter passes through the Sigmoid function again, the spatial attention parameter is dotted with the last half channel features of the corresponding location activation maps of the Re-ID

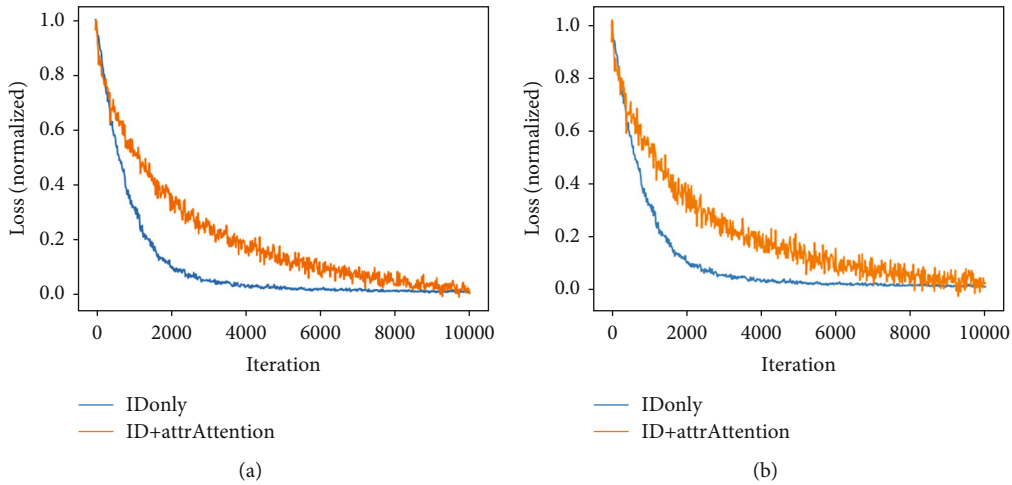

Figure 4: Comparison of training loss of the two network: (a) Market1501; (b) DukeMTMC-reID.

network in blocks to get the final spatial attention (i.e., the feature map contains salient attribute information).

4. Experiments

First of all, we introduced the dataset containing evaluation protocol, loss function, and training details used in Section 4.1; then, we designed a series of ablation experiments in Section 4.2 to verify the effectiveness of our scheme design; finally, in Section 4.3, we compared our model with the advanced Re-ID algorithm to illustrate the superiority of our algorithm.

4.1. The Experimental Details

4.1.1. Dataset

(1) DukeMTMC-reID. DukeMTMC-reID is a person Re-ID subset of DukeMTMC dataset and provides manually annotated bounding boxes. DukeMTMC-reID contains 16,522 training images of 702 of these pedestrian IDs, 2,228 query images from another 702 pedestrian IDs, and 17,661 images' gallery of 702 pedestrian IDs.

(2) Market1501. It was captured on the campus of Tsinghua University in the summer with 6 cameras. The dataset contains a total of 32,688 images with 1,501 pedestrian IDs. Among them, the training set contains 12,936 images of 751 pedestrian IDs, the query set contains 3,368 images of 750 pedestrians, and the test set includes 16,384 images of 750 pedestrians, all of whom have appeared in at least two cameras.

We adopt the Cumulative Matching Characteristics (CMC) at Rank-1 and the mean Average Precision (mAP) as the evaluation indicators to test the performance of different Re-ID methods on these datasets. The mAP is the mean of Average Precision (AP) for each query image. Rank-1 is the probability that the top image in the search results is the target.

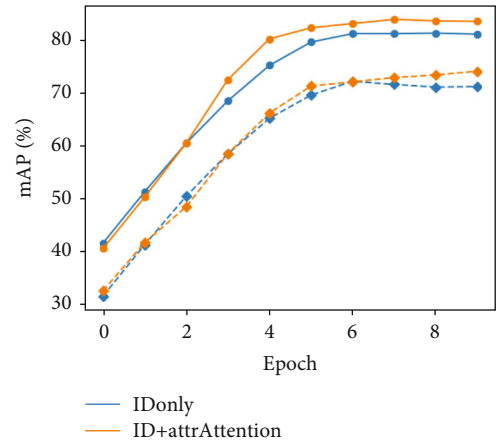

Figure 5: Comparison of mAP change trends of the two networks during training.

4.1.2. Loss Function.

We set both attribute loss and identification loss to cross-entropy loss (in our code, it is the combination of Softmax function and cross-entropy function), and distance loss as triplet loss.

4.1.3. Training Details.

In order to ensure the consistency of the experimental results, the experimental process is carried out in the same software and hardware environment. The experimental platform is based on 64-bit Ubuntu18.04 operating system, the device memory is 32 G, the CPU is Intel® Xeon E5-2678V3 CPU @2.5 GHz, and the training is conducted on the NVIDIA GTX1080TI single GPU platform, the CUDA version is 10.2, and the experimental framework is based on the PyTorch 1.6.0 version.

We set the size of the input pedestrian images as $256 * 128$, and use data augmentation methods such as random erasing and image expansion during the training process. Stochastic Gradient Descent (SGD) is selected by the network parameter optimization algorithm. In particular, we increase the learning rate from $3e$-5 to $1e$-3 in the first training epochs and then decrease it to $5e$-4, $1e$-4, and $3e$-5 in the

Table 2: Performance comparison between our SAN-GAL and other classic Re-ID methods.

Methods	Market1501		DukeMTMC-reID	
	Rank-1 (%)	mAP (%)	Rank-1 (%)	mAP (%)
SVDNet (CVPR17) [24]	82.3	62.1	76.7	56.8
PAN (TCSVT18) [25]	82.2	63.4	71.6	51.5
DuATM (CVPR18) [26]	91.4	76.6	81.8	64.6
PCB (ECCV18) [27]	92.3	77.4	81.8	66.1
SPReID (CVPR18) [28]	92.5	81.3	84.4	70.1
VPM (CVPR19) [29]	93.0	80.8	83.6	72.6
AANet (CVPR19) [29]	93.9	83.4	87.7	74.3
IANet (CVPR19) [30]	94.4	83.1	87.1	73.4
SAN-GAL (ours)	94.4	83.9	85.8	74.1
BFE (ICCV19) [31]	95.3	86.2	88.9	75.9

5th, 7th, and 15th training epochs, respectively, which ends after a total of 20 training epochs.

4.2. Ablation Experiments. In order to fully verify the effectiveness of our proposed module and method, we conduct three ablation experiments on Market1501 and DukeMTMC-reID. The specific experimental differences are as follows: firstly, only the pedestrian global ID information is used to train the Re-ID network as the baseline; then, the pedestrian attribute classification network is added on the basis of baseline. Finally, on the basis of the above steps, the ASAM is introduced to form the final design scheme. The experimental results are shown in Table 1.

As shown in Table 1, the Rank-1 and mAP values of IDonly method on the Market1501 dataset are 92.6% and 81.3%, and the Rank-1 and mAP values on the DukeMTMC-reID dataset are 82.3% and 72.2%. After the introduction of attribute information, compared with IDonly method, the mAP of ID+attr method on Market1501 and DukeMTMC-reID increases by 0.8% and 0.2%, but the Rank-1 in Market1501 decreases by 0.2%, and the Rank-1 in DukeMTMC-reID only increases by 0.1%. It can be seen that simply adding attribute classification network cannot bring better effect to the Re-ID task. After introducing our attention mechanism on the basis of ID+attr method, the improvement is significant, especially the accuracy improvement of Rank-1 on the two datasets is 2% and 3.4%, respectively, which fully proves the effectiveness of our ASAM and the overall design.

During the training process, the ordinates of the loss value of IDonly and ID+attrAttention are normalized as shown in Figure 4 (only part of iteration is intercepted). It can be seen that in the Re-ID network with attribute attention mechanism, although loss declines slowly in the initial stage, it is still in a downward trend when IDonly method converges early and eventually achieves loss value lower than that of IDonly method. The change trend comparison of mAP shown in Figure 5, it also supports the enhancement effect of ID+attrAttention method on the Re-ID task.

4.3. Comparison of Algorithms. To demonstrate the superiority of our SAN-GAL in the overall design, we select some representative algorithms in the Re-ID field for comparison, and the selection principles are as follows:

(1) All of them are all based on convolutional neural networks

(2) All of them are representative algorithms in different Re-ID genres

(3) Experiments were carried out on Market1501 and DukeMTMC-reID datasets and evaluated by Rank-1 and mAP

As shown in Table 2, our SAN-GAL algorithm outperforms most of the algorithms on both Market1501 and DukeMTMC-reID datasets. Among all the algorithms listed, compared with PAN in 2018 TCSVT, the Rank-1 and mAP of our SAN-GAL on Market1501 are improved by 12.2% and 20.5%, respectively, and the Rank-1 and mAP of our SAN-GAL on DukeMTMC-reID are improved by 14.2% and 22.6%, respectively. Compared with the VPM in 2019, SAN-GAL is 3.1% higher than its mAP on Market1501. However, compared with BFE, there is still a 0.9% gap in Rank-1 on Market1501. In particular, our SAN-GAL does not add any feature enhancement module or special training technique other than the introduction of a specific attribute attention mechanism. Therefore, our SAN-GAL is an algorithm with both performance and potential.

5. Conclusion

In order to overcome the limitation of global pedestrian features in cross-device scenarios, we proposed SAN-GAL. Different from the previous approach of simply increasing the branch of attribute binary classification network, our SAN-GAL network is guided by attribute labels. Firstly, by generating AAH, we can accurately locate the fine-grained attributes of pedestrians. Then, on the basis of AAH, ASAM is constructed to integrate the global ID information and local attribute information to enhance the discrimination. In particular, our dual-trace network does not need additional attribute region annotation on the dataset, so it has better flexibility and adaptability. By testing on Market1501

and DukeMTMC-reID datasets, the effectiveness and superiority of our scheme design are proved. In the future, we want to expand and apply these ideas to other computer vision tasks, such as Person Search (PS).

Data Availability

Previously reported DukeMTMC-reID and Market1501 data were used to support this study and are available at 10.1109/ICCV.2017.405 and 10.1109/ICCV.2015.133, respectively. These prior studies (and datasets) are cited, respectively, at relevant places within the text as references [19, 20].

Conflicts of Interest

Center for Public Security Information and Equipment Integration Technology agreed to publish this paper, and all authors declare that they have no conflict of interest.

Acknowledgments

This research was supported by the Center for Public Security Information and Equipment Integration Technology of UESTC (University of Electronic Science and Technology of China). Especially, thanks to the computing platform provided by the laboratory.

References

[1] Z. Cai and Z. Xu, "A private and efficient mechanism for data uploading in smart cyber-physical systems," *IEEE Transactions on Network Science and Engineering (TNSE)*, vol. 7, no. 2, pp. 766–775, 2020.

[2] Z. Xu and Z. Cai, "Privacy-preserved data sharing towards multiple parties in industrial IoTs," *IEEE Journal on Selected Areas in Communications (JSAC)*, vol. 38, no. 5, pp. 968–979, 2020.

[3] Z. Cai and Z. He, "Trading private range counting over big IoT data," in *The 39th IEEE International Conference on Distributed Computing Systems (ICDCS 2019)*, Dallas, TX, USA, 2019.

[4] W. S. Zheng, S. Gong, and T. Xiang, "Reidentification by relative distance comparison," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 3, pp. 653–668, 2013.

[5] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, vol. 1, pp. 886–893, San Diego, CA, USA, 2005.

[6] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the seventh IEEE international conference on computer vision*, vol. 2, pp. 1150–1157, Kerkyra, Greece, 1999.

[7] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Deep metric learning for person re-identification," in *2014 22nd international conference on pattern recognition*, pp. 34–39, Stockholm, Sweden, 2014.

[8] W. Li, R. Zhao, T. Xiao, and X. Wang, "Deepreid: deep filter pairing neural network for person re-identification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 152–159, Columbus, USA, 2014.

[9] A. Hermans, L. Beyer, and B. Leibe, "In defense of the triplet loss for person re-identification," 2017, https://arxiv.org/abs/1703.07737.

[10] W. Chen, X. Chen, J. Zhang, and K. Huang, "Beyond triplet loss: a deep quadruplet network for person re-identification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 403–412, Honolulu, USA, 2017.

[11] D. Chen, D. Xu, H. Li, N. Sebe, and X. Wang, "Group consistent similarity learning via deep CRF for person re-identification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 8649–8658, Salt Lake City, USA, 2018.

[12] M. Geng, Y. Wang, T. Xiang, and Y. Tian, "Deep transfer learning for person re-identification," 2016, https://arxiv.org/abs/1611.05244.

[13] P. Fang, J. Zhou, S. K. Roy, L. Petersson, and M. Harandi, "Bilinear attention networks for person retrieval," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 8030–8039, Seoul, Korea (South), 2019.

[14] Y. Fu, X. Wang, Y. Wei, and T. Huang, "STA: Spatial-temporal attention for large-scale video-based person re-identification," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, pp. 8287–8294, Hawaii, USA, 2019.

[15] S. Li, S. Bak, P. Carr, and X. Wang, "Diversity regularized spatiotemporal attention for video-based person re-identification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 369–378, Salt Lake City, USA, 2018.

[16] Y. Lin, L. Zheng, Z. Zheng et al., "Improving person re-identification by attribute and identity learning," *Pattern Recognition*, vol. 95, pp. 151–161, 2019.

[17] Z. Cai, Z. He, X. Guan, and Y. Li, "Collective data-sanitization for preventing sensitive information inference attacks in social networks," *IEEE Transactions on Dependable and Secure Computing*, vol. 15, no. 4, pp. 577–590, 2018.

[18] Z. Cai, Z. Xiong, H. Xu, P. Wang, W. Li, and Y. Pan, "Generative adversarial networks," *ACM Computing Surveys*, vol. 54, no. 6, pp. 1–38, 2021.

[19] Z. Zheng, L. Zheng, and Y. Yang, "Unlabeled samples generated by GAN improve the person re-identification baseline in vitro," in *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 3774–3782, Venice, Italy, 2017.

[20] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable person re-identification: a benchmark," in *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 1116–1124, Santiago, Chile, 2015.

[21] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, pp. 618–626, Venice, Italy, 2017.

[22] X. Liu, H. Zhao, M. Tian et al., "Hydraplus-net: attentive deep features for pedestrian analysis," in *Proceedings of the IEEE international conference on computer vision*, pp. 350–359, Venice, Italy, 2017.

[23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, Las Vegas, USA, 2016.

[24] Y. Sun, L. Zheng, W. Deng, and S. Wang, "Svdnet for pedestrian retrieval," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3800–3808, Venice, Italy, 2017.

[25] Z. Zheng, L. Zheng, and Y. Yang, "Pedestrian alignment network for large-scale person re-identification," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 10, pp. 3037–3045, 2018.

[26] J. Si, H. Zhang, C. G. Li et al., "Dual attention matching network for context-aware feature sequence based person re-identification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5363–5372, Salt Lake City, USA, 2018.

[27] Y. Sun, L. Zheng, Y. Yang, Q. Tian, and S. Wang, "Beyond part models: person retrieval with refined part pooling (and a strong convolutional baseline)," in *Proceedings of the European conference on computer vision (ECCV)*, pp. 480–496, Munich, Germany, 2018.

[28] M. M. Kalayeh, E. Basaran, M. Gökmen, M. E. Kamasak, and M. Shah, "Human semantic parsing for person re-identification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1062–1071, Salt Lake City, USA, 2018.

[29] C. P. Tay, S. Roy, and K. H. Yap, "Aanet: attribute attention network for person re-identifications," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7134–7143, Long Beach, USA, 2019.

[30] R. Hou, B. Ma, H. Chang, X. Gu, S. Shan, and X. Chen, "Interaction-and-aggregation network for person re-identification," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9317–9326, 2019.

[31] Z. Dai, M. Chen, X. Gu, S. Zhu, and P. Tan, "Batch dropblock network for person re-identification and beyond," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3691–3701, 2019.