IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 21 March 2024, accepted 8 April 2024, date of publication 11 April 2024, date of current version 22 April 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3387478

RESEARCH ARTICLE

Cross-Domain Person Re-Identification Based on Normalized IBN-Net

XUEMEI BAI[1], AO WANG[1], CHENJIE ZHANG[1], AND HANPING HU[2]
[1]School of Electronic Information Engineering, Changchun University of Science and Technology, Changchun 130000, China
[2]School of Computer Science and Technology, Changchun University of Science and Technology, Changchun 130000, China

Corresponding author: Chenjie Zhang (zhangcj@cust.edu.cn)

This work was supported by the Science and Technology Development Programme of Jilin Province, under Project 20240302089GX.

ABSTRACT In existing methods, the data distribution between different domains may have large differences, which can lead to performance degradation in the target domain, and the modelling of feature variability between different domains is not sufficient. Aiming at the problem of severe performance degradation during cross-domain migration, this paper proposes a cross-domain person re-identification method based on normalized IBN-Net. The normalized IBN-Net network introduces instance normalization and batch normalization to handle the feature maps at different scales. First, this study employed the normalized IBN-Net network as the backbone of the ResNet50 network. Second, the SimAM attention mechanism is integrated into the backbone network, which is an attention mechanism for inter-modal fusion,that is mainly used in multimodal data processing tasks,and it learns the spatial attention weights in the person images to obtain person information with more discriminative features. Finally, supervised learning was performed using the cross-entropy loss function during source domain training. Meanwhile, it can obtain the details of the source domain samples using the triplet loss function, thereby improving the classification performance. During target domain training, adaptation to the challenges of viewpoint, illumination, background, and feature distribution differences between samples is achieved by learning variations between the source and target domains. In the test stage, comparative experiments were conducted on two large-scale public data sets, Market-1501 and DukeMTMC-reID, achieving Rank-1 accuracies of 86.6% and 79.3%, respectively, with mean average precision mAPs of 68.7% and 62.6%, respectively. The experimental results show that the proposed method performs better in terms of improving the generalization ability of the model.

INDEX TERMS Cross-domain, person re-identification, IBN-Net, attention mechanism, normalization.

I. INTRODUCTION

Pedestrian re-identification [1], [2], [3] utilizes computer vision techniques aimed at extracting and matching pedestrian features in images and videos. This technique achieves accurate identification and tracking of pedestrian identity in different scenarios. It is widely used in the fields of intelligent security, intelligent transportation, and video surveillance. With the promotion and application of deep learning techniques, the accuracy of supervised learning [4] in the field of person re-identification has significantly improved. However, supervised learning has several limitations. In other words, it relies on datasets with real labels and requires a datasets to provide supervised training samples. Therefore the model can learn the data;however,it is difficult to meet the demands of practical applications.

Owing to the differences in camera parameters, background, and illumination in different scenes [5], [6], models trained on the source domain are directly applied to the target domain datasets, which can lead to the degradation of person re-identification performance. In addition, traditional cross-domain problems typically assume that the source and target domains have exactly the same categories. In the pedestrian re-identification problem, This assumption does not apply. Cross-domain pedestrian re-identification datasets are typically collected at different times and space. Images from source and target domains often contain different types of identity information. which makes cross-domain person re-identification should be regarded as an open-set problem, which is more challenging compared to the

The associate editor coordinating the review of this manuscript and approving it for publication was Yiqi Liu.

closed-set problem. Therefore, when solving this problem, we must consider these factors and adopt corresponding strategies to improve the performance of person re-identification.

In this paper, we study cross-domain methods, focusing on solving the problem of large feature differences between different datasets in existing cross-domain person re-identification methods, and propose a network model based on normalized IBN-Net with the following main innovations and contributions.

1) In this study, we propose a network model based on a normalized IBN-Net, which combines instance normalization and batch normalization network structures to solve feature disparity problems in cross-domain image classification tasks. By entirely using the appearance and structural features, the normalized IBN-Net network can improve the model's generalization performance and has a better fitness to image data from different domains.

2) In this study, we proposed a SimAM attention mechanism fusion strategy. The SimAM attention mechanism performs weighted fusion of input features through similarity metrics and attention weight calculations to obtain more informative and discriminative feature representations.which can automatically learn the critical features and suppress irrelevant or redundant features in the task. This ability to selectively enhance the feature representation helps improve the network's discriminative and generalization capabilities, resulting in better performance of the normalized IBN-Net network in tasks such as image classification.

3) Compared with the algorithms proposed in recent years, the cross-domain recognition accuracy of the proposed algorithm was improved on two public datasets, DukeMTMC-ReID and Market1501.

II. RELATED WORK

The research on unsupervised cross-domain person re-identification was developed based on unsupervised person re-identification. Compared with the unsupervised approach, the unsupervised cross-domain system utilizes a labeled source domain datasets, which provides the researcher with a priori knowledge to better guide the re-recognition task, leading to better results.

The current cross-domain person re-identification methods can be mainly divided into five main categories.

A. CROSS-DOMAIN FEATURE ALIGNMENT [7], [8]

The domain gap was reduced by aligning the data distributions in the source and target domains. Aligning the data distribution in the target domain makes the data distribution in the target domain as consistent as possible with the data distribution in the source domain, which can reduce the domain gap and thus improve the accuracy of the experiment. Wang et al. [7] used the additional labelled pedestrian attribute information to train the network by combining the identity label branch and the attribute branch to achieve information intermingling between the networks, and finally learn the more essential features of the pedestrians.

B. CROSS-DOMAIN IMAGE GENERATION

This method uses generative adversarial networks (GAN) to process data images to obtain similar image styles between the datasets. Liu et al. [9] decomposed the cross-domain transformation into three factor transformations, namely, illumination, camera angle and resolution, where each factor was treated as a sub-style, trained a generator for each sub-style, and proposed an adaptive-transfer network (ATNet) that can weigh the influence of various factors and thus carry out the fusion process. They achieved fine-grained style migration by minimising the sub-tasks at the intermediate layer. The network achieves fine-grained style migration by minimizing sub-tasks in the middle layer.

C. CLUSTERING-BASED APPROACHES

To make full use of unlabelled target domain data, pseudo-labels generated by clustering algorithms are used as labels of the target domain, and this type of method has been proven to have the best effect at present in a large number of experiments. Fu et al. [10] proposed a self-similarity grouping (SSG) model by vertically averaging the feature map into six local features, and using local features to assign multi-scale clustering pseudo-labels. Zhai et al. [11] proposed a new discriminative clustering method, augmented discriminative clustering (AD-Cluster), to solve the problem of unsupervised cross-domain pedestrian re-identification through a density-based clustering algorithm with adaptive sample expansion and discriminative feature learning.

D. CALCULATING THE SOFT LABELS GENERATED BY FEATURE SIMILARITY TO OPTIMISE THE NETWORK

Usually soft labels are represented using either the average features saved across all images or a feature space constructed using an auxiliary datasets. Feature similarity is calculated by constructing a feature space to represent the images to be trained using either already trained images or auxiliary datasets images.The exemplar camera neighborhood invariance (ECN) proposed by Zhong et al. [12] uses the memory structure to save the average features to assign soft labels to the training images, and supervises the optimization of the network using three invariants: sample invariance, camera invariance, and neighborhood invariance.

In practical applications, cross-domain feature alignment and cross-domain image generation methods usually requires complex calculation and data generation processes, leading to the loss of some feature information. Therefore, this study designed a cross-domain pedestrian reconstruction method based on the normalized IBN-Net [13]. The recognition model can effectively solve the problem of significant feature differences between different datasets in cross-domain person re-identification methods. This representational and generalization capabilities.normalization approach differs from

traditional normalization methods in that it is specifically designed for pedestrian re-identification task.

This model combines the network structure of instance normalization and batch normalization three times, integrates the SimAM attention mechanism [14], and uses a fused network structure for training to improve the model's adaptability to changes in image appearance. The attention mechanism can dynamically allocate attention weights to different areas based on the input image content,thereby allowing the network to pay more attention to important information in the image. By integrating the SimAM attention mechanism into IBN-Net, the network can better capture useful features in images and improve the expression ability of features. In future research directions,the relationship between different types of normalization methods and pedestrian feature representations can be further explored, and more effective normalization mechanisms can be designed to enhance model performance.

III. CROSS-DOMAIN PERSON RE-IDENTIFICATION MODEL BASED ON NORMALIZED IBN-NET AND FUSED WITH ATTENTION MECHANISM

A. PROBLEM DEFINITION

For cross-domain person re-identification tasks, there are usually two data domains:the source domain $S=\{X_s, Y_s\}$, which represents the image in the source domain, and x represents the corresponding real label. In addition, there is no label information in the target domain $T=\{X_t\}$,and only the target domain is imaginary. This study aims to extract more discriminative pedestrian features to learn changes in the target domain, thereby improving the re-identification performance of the model when transferring from the source to the target domain.

B. NORMALIZED IBN-NET NETWORK

The feature difference between different data sets is an important factor that affect the performance degradation of cross-domain person re-identification. When there is a significant difference in appearance between the training and test data, the performance of the model decreases significantly, which is the difference between different domains. For example, the target light in the training data is intense, and the target light in the test data is dim;therefore, the general effect is not very good. This study used the normalized IBN-Net network to reduce domain difference.

The normalized IBN-Net uses a three-time combination of Instance Normalization and Batch Normalization.

The first step is to combine the instance normalization layer and batch normalization layer in series in the backbone network of the ResNet50 network. Using instance normalization in the backbone network helps normalize the entire feature map, thereby alleviating vanishing gradient or exploding gradient problems. The training stability of the network can be improved, making it easier to train, and instance normalization helps maintain the relative relationship between features, thus improving the expressive ability of the features, which is very beneficial for feature learning in backbone networks, as these features often need to capture more global information.

The input feature map is denoted as X,which is a four-dimensional tensor with a shape of (n, c, h, w). N represents the batch size, c represents the number of channels, H represents the height, and W represents the width. For each sample (n) and channel (c), the mean and variance of the channel in the sample are calculated.

$$\mu_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} X_{n,c,i,j} \tag{1}$$

$$\sigma_c^2 = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} (X_{n,c,i,j} - \mu_c) \tag{2}$$

Normalize each channel using the calculated mean and variance.

$$\tilde{X}_{n,c,i,j} = \frac{X_{n,c,i,j} - \mu_c}{\sqrt{\sigma_c^2 + \rho}} \tag{3}$$

where, ρ is a small positive number, usually used to prevent variance from 0. This study used learnable parameters γ_c, β_c to scale and translate the normalized features.

$$Y_{n,c,i,j} = \gamma_c \tilde{X}_{n,c,i,j} + \beta_c \tag{4}$$

Finally, the output Y processed by the instance normalization layer is obtained, and the shape of Y is the same as that of the input X, that is, (N, C, H, W).

The second step was to combine instance normalization and batch normalization in parallel after the conv1 layer under the bottleneck layer in each layer of the ResNet50 network. Given the input x, the output of the bottleneck layer is F(x), the result of instance normalization is IN(F(X)), and the result of batch normalization is BN(F(X)), then the process of combining the two can be expressed as:

$$y = \alpha IN(F(x)) + \beta BN(F(x)) \tag{5}$$

where, α and β are learnable weight parameters used to adjust the relative weight of instance normalization and batch normalization,respectively.This combination can enhance the expressive ability of the network and enable it to normalize data at different levels, thereby achieving a more flexible architecture design.

The third combination combined instance normalization and batch normalization in series after the conv3 layer under the bottleneck layer in each layer of the ResNet50 network. For instance normalization, the formula is expressed as:

$$IN(x^i) = \frac{x^i - \mu^i}{\sqrt{\sigma^i + \varepsilon}} \tag{6}$$

where, x^i represents the mean value of the i-th layer feature map, σ^i represents the variance of the i-th layer feature map, and ε is a small constant used for stable calculations. The formula for batch normalization is expressed as:

$$BN(x^i) = \omega^i \frac{x^i - \mu^i}{\sqrt{\sigma^i + \varepsilon}} + \theta^i \tag{7}$$

where, ω^i and θ^i are learnable parameters that scale and translate the normalized feature map. By combining instance normalization and batch normalization in series, a new feature map can be obtained, denoted as Y^i, where i represents the number of layers of the network, which is expressed as:

$$Y^i = BN(IN(x^i)) \tag{8}$$

where, σ and φ are learnable weight parameters used to adjust the relative weights of the instance normalization and batch normalization. This series combination method can simultaneously reduce the deviation between small batches and the statistical difference in features, thereby improving the network's convergence speed, training stability, and generalization ability. The normalized IBN-Net is illustrated in Fig.1.

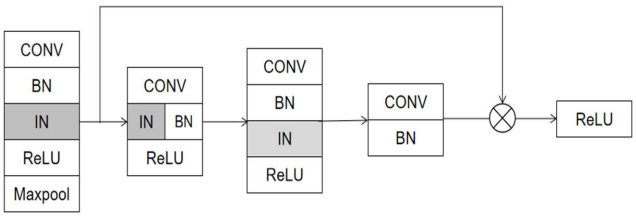

FIGURE 1. Structure of normalized IBN-Net network.

C. SimAM ATTENTION MECHANISM

SimAM (three-dimensional attention mechanism) is a parameter-free self-attention module similar to the human brain. There are abundant neurons in the brain. To mine more important neurons, an energy function is constructed to determine each importance of neurons.In neuroscience, information-rich neurons often exhibit different firing patterns from surrounding neurons,leading to the phenomenon of spatial inhibition, in which activated neurons inhibit other surrounding inactive neurons. Therefore, neurons with spatial inhibitory effects should be given greater importance. Thus, through the SimAM self-attention mechanism, important features on the image can be provided with higher weights, thereby making the network pay more attention to this aspect. A schematic diagram of the SimAM attention mechanism is shown in Fig.2.

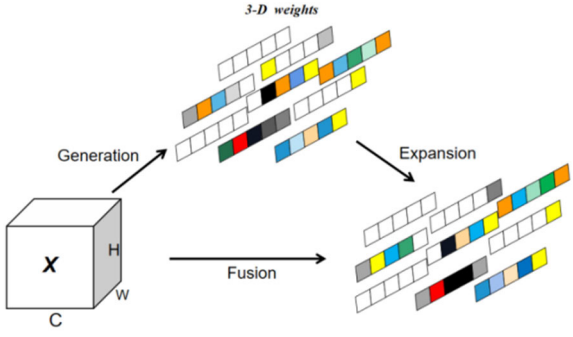

FIGURE 2. Structure of SimAM attention mechanism.

Important neurons are identified by measuring the linear separability between neurons,;therefore, the following energy function is defined.

where, $\hat{t} = w_t t + b_t$, $\hat{x_i} = w_t x_i + b_t$, are the linear transformations of t and x_i respectively, t and x_i are the target neuron and other neurons in a single channel of the input feature, i is the spatial index, and $M = H \times W$ is the number of neurons in the channel, w_t is the weight and b_t is the bias. All values in Equation (9) are scalars, and Equation (9) reaches its minimum when $t = y_t$ and x_i are y_0 both. y_t and y_0 are two different values. By minimizing this equation, Equation (9) is equivalent to finding the linear separability between the target neuron t and all other neurons in the same channel. For simplicity, we adopt binary labels (1 and -1) for y_t and y_0 and add a regularizer in equation (10); λ is the regularization parameter, and λw_t^2 is the regularization term. The final energy function is defined as follows:

$$e_t(w_t, b_t, y, x_i) = \left(y_t - \hat{t}\right)^2 + \frac{1}{M-1}\sum_{i=1}^{M-1}(y_0 - \hat{x}_i)^2 \tag{9}$$

$$e_t(w_t, b_t, y, x_i) = \frac{1}{M-1}\sum_{i=1}^{M-1}(-1-(w_t x_i + b_t))^2 + (1-(w_t x_i + b_t))^2 + \lambda w_t^2 \tag{10}$$

The analytical solution of the above formula is.

$$w_t = -\frac{2(t-\mu_t)}{(t-\mu_t)^2 + 2\sigma_t^2 + 2\lambda} \tag{11}$$

$$b_t = -\frac{1}{2}(t+\mu_t)w_t \tag{12}$$

Because all the neurons in each channel follow the same distribution, we can first calculate the mean and variance of the input features in the H and W dimensions to avoid repeated calculations.

$$e_t^* = \frac{4\left(\hat{\sigma}^2 + \lambda\right)}{\left(t-\hat{\mu}\right)^2 + 2\hat{\sigma}^2 + 2\lambda} \tag{13}$$

where, $\hat{\mu}$ and $\hat{\sigma^2}$ are the mean and variance of the pixels in a single channel of the feature map, $\hat{\mu} = \frac{1}{M}\sum_{i=1}^{M} x_i$, $\hat{\sigma^2} = \frac{1}{M}\sum_{i=1}^{M}(x_i - \mu)^2$. t is the target neuron in the input channel, and λ is the regularization constant. It can be seen from formula (13) that the lower the energy, the greater the difference between neuron t and the surrounding neurons, and the higher the importance. Therefore, the importance of a neuron can be obtained using $\frac{1}{e_t^*}$. After deriving the energy, the characteristics are enhanced through a sigmoid function, and the output of SimAM is obtained as formula (14).

$$\tilde{X} = \text{Sigmoid}\left(\frac{1}{E}\right) \odot X \tag{14}$$

where, $E = e^*$, X is the input SimAM and $\tilde{X}$ is the output of SimAM.

SimAM (Similarity-based attention mechanism) focuses on the similarities between different parts of the data, which makes it perform well when processing data where

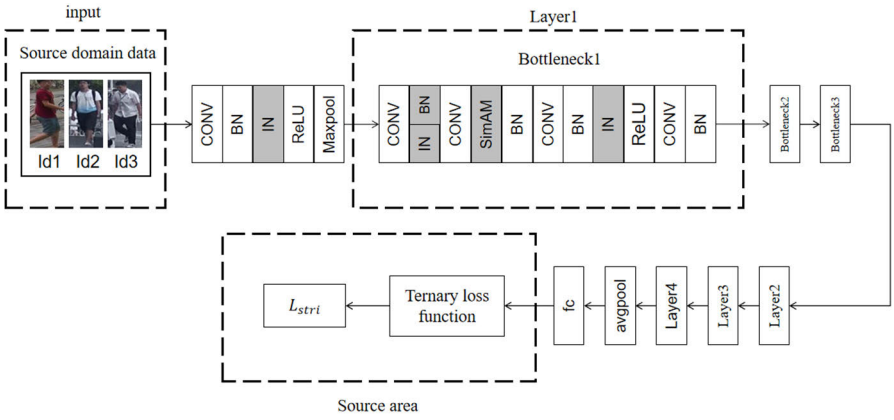

FIGURE 3. The overall structure of the method.

similarities exist, such as face recognition or person re-identification.SimAM helps the model capture important information in the input better by comparing the similarities between features and assigning corresponding weights, thereby improving the representation quality of the features. In addition, SimAM can also help the model ignore irrelevant areas or features in complex scenes or the presence of background interference, thereby improving the robustness of the model. This gives SimAM significant advantages over channel and spatial attention mechanisms in specific tasks and data situations.

D. SimAM ATTENTION MECHANISM

For the cross-domain person re-identification task, the model used ResNet50 as the basic architecture,and the residual block of ResNet50 added a normalized IBN-Net network. This study integrated the SimAM attention mechanism after the conv2 layer under the bottleneck layer in each layer of the ResNet50 network. The overall structure of the proposed method is shown in Fig.3.

The input image was passed to the normalized IBN-Net network. The image first passes through a series of convolutional and pooling layers. These layers extract low-level features, including basic image information such as edges and textures. The normalized IBN-Net network adjusts the mean and variance of the feature map by alternately using the IN layer (Instance Normalization) and the BN layer (Batch Normalization). The IN layer is used to normalize the feature map to reduce redundancy and correlation between features. The BN layer was used to normalize the feature map to ensure that it had an appropriate mean and variance. After normalization, the features were divided into two branches: channel attention branch and spatial attention branch. The channel attention branch mainly focuses on the correlation between the feature channels. In this branch, the SimAM attention mechanism was used to calculated the importance weight of each feature channel, and the spatial attention module calculates the importance weight of each position by performing convolution and pooling operations on the spatial location of the feature. After the feature extraction is completed, a global average pooling operation is performed on the output of the last convolutional layer.

A network model based on the normalized IBN-Net and fused with an attention mechanism combines the normalization mechanism with residual connectivity and the fusion of two branch networks in pedestrian re-identification. Optimization and fusion of features are achieved through interaction, thus improving the network's representational and generalization capabilities. This combination and design can capture the subtle features of pedestrian images better and improve the performance of the model while reducing the risk of over fitting, which is significantly innovative and superior in the pedestrian re-identification task.

E. SOURCE DOMAIN PRE-TRAINING

To transfer the knowledge obtained from the source domain to the target domain more effectively, the model is pre-trained on the source domain data set, and the pre-trained model parameters are used as initialization parameters to train the target domain datasets. Using this strategy, common feature representations are extracted from the source domain data, and the convergence of the model is accelerated in the target domain tasks to achieve better cross-domain performance transfer.

In the context of a deep neural network model M_s with parameter θ, we first perform supervised pre-training on the model and use the source domain data set for training. In the source domain data set, each pedestrian image $x_{s,i} \in X_s$ will be processed by the model to obtain the corresponding feature representation $f(x_{s,i}\theta)$, which will be used for the final identity prediction $p(x_{s,i}|\theta)$. During the model training,the cross-entropy loss and triplet loss were combined. The cross-entropy loss function minimizes the difference between the model's identity prediction output and actual labels, allowing the model to classify pedestrian images more accurately. The cross-entropy loss is defined as formula (15).

$$L_{s,id} = -\frac{1}{N_s}\sum_{i=1}^{N_s} p(y_{s,i}|x_i^{\Theta}) \tag{15}$$

The triplet loss function helps closely cluster image samples with the same identity in the feature space. By contrast, image samples with different identities were pushed apart, enhancing the distinction between features. The triplet loss is defined as formula (16).

$$L_{s,tri} = \frac{1}{N_s} \times \sum_{i=1}^{N_s} \ln(m + ||f_s^+ (x_s^\Theta, i)||_2 - ||f_s^- (x_s^\Theta, i)||_2) \tag{16}$$

By adjusting the parameters of Model A, it can better learn the feature representations of pedestrian identities from the source domain data. Thus, after the source domain pre-training is completed, our model will have better feature extraction capabilities and identity prediction accuracy, laying a good foundation for transfer learning in subsequent tasks.

IV. ANALYSIS OF EXPERIMENTAL RESULTS

To verify the effectiveness of the method, two public pedestrian re-identification data sets,Market1501 and DukeMTMC-ReID. The method proposed in this study was evaluated on these two data sets and compared with the mainstream methods.In addition, ablation experiments and parameter analyses were performed.

A. DATA SETS AND EVALUATION INDICATORS

Pedestrian images in the Market1501 datasets were collected from six cameras on the Tsinghua University campus, and 1,501 pedestrians were annotated. It contained 12,936 pictures of 751 pedestrians for the training set and 19,732 images of 750 pedestrians for the test set. There were no duplicate pedestrian IDs in the training and the test sets, which means that 751 of them appeared in the training set. No pedestrians appeared in the test set.

The DukeMTMC datasets was collected from eight cameras at Duke University. The datasets was stored in videos containing manually annotated pedestrian bounding boxes. The DukeMTMC-reID data set collected an image every 120 frames from the video of the DukeMTMC data set to form the DukeMTMC-reID data set. The DukeMTMC-reID datasets contains 36,411 images of 1,812 pedestrians.A total of 1,404 pedestrians were captured by more than two cameras, whereas 408 pedestrians were captured by only one camera. Among these, 16,522 images containing 702 pedestrians were used as the training set, and 17,661 images containing 702 pedestrians and 408 interference pedestrians were used as the testing set.

In this study, we used the mean average precision (mAP) and ranking accuracy (Rank-n Accuracy) to quantitatively evaluate the performance of the pedestrian re-identification model. mAP is obtained by taking a comprehensive weighted average of the average accuracy of all categories, while Rank-n accuracy evaluates the accuracy among the top n candidates in the retrieval results. This study mainly evaluates Rank-1, Rank-5, and Rank-10.

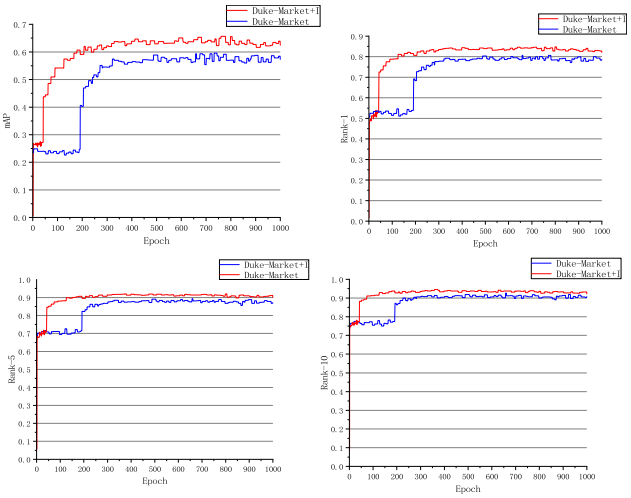

FIGURE 4. Duke-Market normalized IBN-Net network module evaluation experiment.

B. DATA SETS AND EVALUATION INDICATORS

The model's training in this article includes pre-training in the source domain and cross-domain adaptation in the target domain. Before the image is input to the network, the size of the image is adjusted to 256 × 128.

The experiment in this article is based on the Pytorch framework and uses the Pytorch1.13.1 version. Use one TITANXpGPU for training and one TITANXpGPU for testing.

C. DATA SETS AND EVALUATION INDICATORS

To verify the effect of each module on model performance, this paper conducted corresponding ablation experiments. The experimental results of each module are shown in Table.1. Among them, B represents the baseline, Table.1. represents the normalized IBN-Net network module, and S represents the SimAM attention mechanism. After adding the normalized IBN-Net network module, the rank-1, rank-5, rank-10, and mAP of migrating from DukeMTMC-reID to Market-1501 increased by 4.1%, 2.7%, 2%, and 6.2% respectively. As shown in Fig.4, the rank-1, rank-5, rank-10, and mAP of migrating from Market-1501 to DukeMTMC-reID increased by 0.1%, 1.2%, 0.3%, and 0.8%, as shown in Fig.5. IBN-Net module By introducing Instance Normalization and Batch Normalisation in the network. This normalization operation helps the model to better capture the detailed information in the image and reduces the correlation between the features, which improves the performance of the model in cross-domain pedestrian re-identification tasks.

After integrating the SimAM attention mechanism module into the residual block of the normalized IBN-Net network, the model effect has also been greatly improved, migrating from DukeMTMC-reID to rank-1, rank-5, and rank-10 of Market-1501 and mAP increased by 7%, 5.4%, 3.9%, and 2.6% respectively, as shown in Fig.6; rank-1, rank-5, rank-10, and mAP that migrated from Market-1501 to DukeMTMC-reID increased by 0.2 %, 0.3%, 0.1%, 0.7%,

TABLE 1. Effect % of different modules.

Module	DukeMTMC-Market1501				Market1501-DukeMTMC			
	Rank-1	Rank-5	Rank-10	mAP	Rank-1	Rank-5	Rank-10	mAP
B	80.7	89.4	92.6	59.8	77.6	85.5	87.8	60.3
B+I	84.8	92.1	94.6	66.0	77.7	86.7	88.1	61.1
B+S	86.1	93.3	95.2	66.8	77.8	85.8	87.9	61.0
B+S+I（Ours）	86.6	93.8	95.6	68.7	79.3	86.4	88.8	62.6

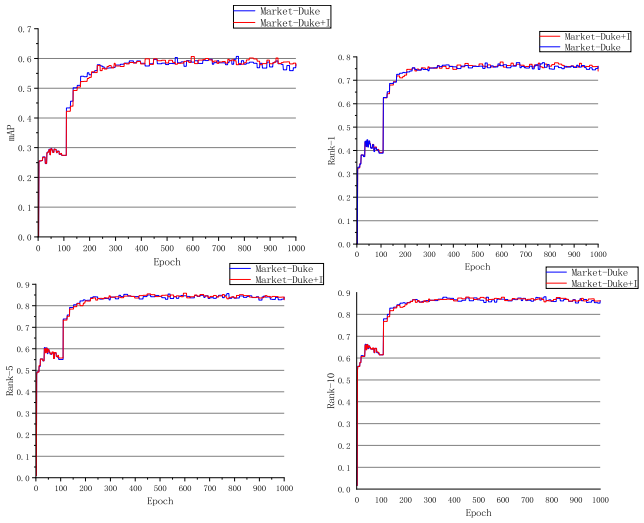

FIGURE 5. Market-Duke normalized IBN-Net network module evaluation experiment.

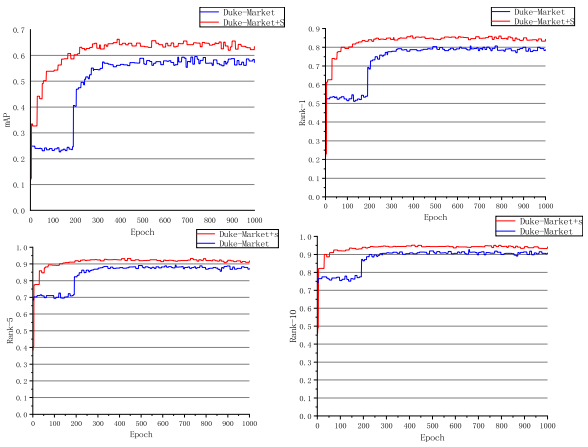

FIGURE 6. Duke-Market SimAM attention mechanism module evaluation experiment.

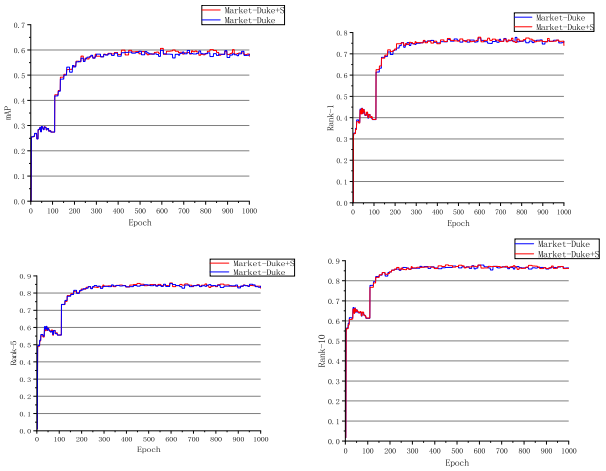

FIGURE 7. Market-Duke SimAM attention mechanism module evaluation experiment.

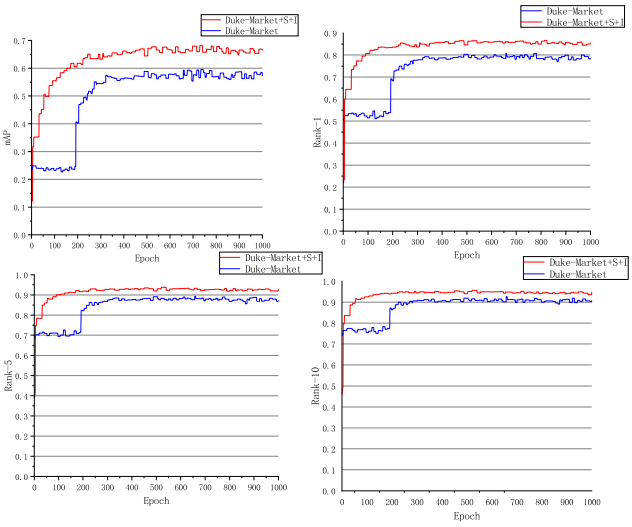

FIGURE 8. Duke-Market SimAM module and normalized IBN-Net module evaluation experiment.

as shown in Fig.7. SimAm attention mechanism introduces spatial and intensity information in the attention mechanism that can help the model to reduce redundant information in the image. This improves the model's performance.

Finally, after adding the SimAM attention mechanism and normalized IBN-Net network module at the same time, the rank-1, rank-5, rank-10, and mAP of migrating from DukeMTMC-reID to Market-1501 increased by 5.9%, 4.4%, 3%, 8.9%, as shown in Fig.8; the rank-1, rank-5, rank-10 and mAP of migrating from Market-1501 to DukeMTMC-reID increased by 1.7%, 0.9%, 1% and 2.3% respectively. As shown in Fig.9. The SimAm attention mechanism can help the model to automatically learn the relationship between global and local features in an image, while IBN-Net can fully consider the importance of global and local features in the process of feature fusion, enabling the model to more comprehensively characterize pedestrian images in cross-domain pedestrian re-identification tasks.

TABLE 2. Comparison with current advanced methods on Market1501 and DukeMTMC-reID data sets.

Method	DukeMTMC-Market1501				Market1501-DukeMTMC			
	mAP	Rank-1	Rank-5	Rank-10	mAP	Rank-1	Rank-5	Rank-10
BUC	38.3	66.2	79.6	84.5	27.5	47.4	62.6	68.4
SSL	37.8	71.7	87.4	37.8	28.6	52.5	63.5	68.0
MMFA	27.4	56.7	75.0	81.8	24.7	45.3	59.8	66.3
TJ-AIDL	26.5	58.2	74.8	81.1	23.0	44.3	59.6	65.0
TAL-MIRN	40.0	73.1	86.3	--	41.3	63.5	76.7	--
ATNet	25.6	55.7	73.2	79.4	24.9	45.1	59.5	64.2
SPGAN+LMP	26.7	57.7	75.8	82.4	26.2	46.4	62.3	68.0
HHL	31.4	62.2	78.8	84.0	27.2	46.0	61.0	66.7
ECN	43.0	75.1	87.6	91.6	40.4	63.3	75.8	80.4
UDAP	53.7	75.8	80.5	93.2	49.0	68.4	80.1	83.5
PCB-PAST	54.6	78.4	--	--	54.3	72.4	--	--
SSG	58.3	80.0	90.0	92.4	53.4	73.0	80.6	83.2
PREST	62.4	82.5	92.1	94.9	56.1	74.4	83.7	85.9
SILC	61.8	80.7	90.1	93.0	50.3	68.5	80.2	85.4
Proposed method	68.7	86.6	93.7	95.6	62.6	79.4	86.4	88.8

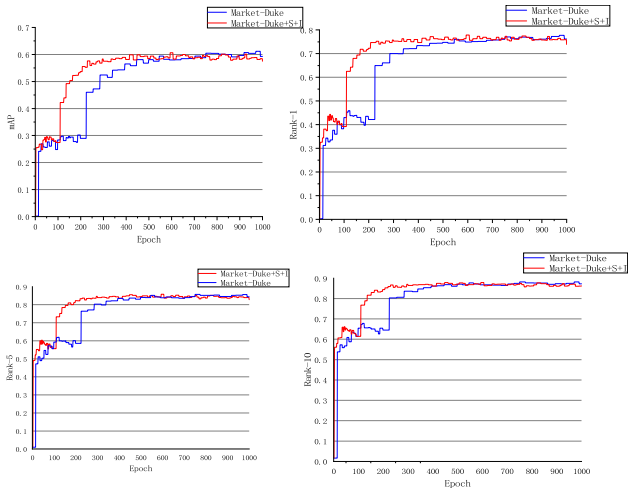

FIGURE 9. Evaluation experiment of Market-Duke SimAM module and normalized IBN-Net module.

D. DATA SETS AND EVALUATION INDICATORS

A comparative experimental verification of cross-domain person re-identification performance was carried out on the Market-1501 and DukeMTMC-reID data sets with the current advanced models to verify the effectiveness of the model proposed in this article. The comparison methods include 1) Unsupervised method, bottom-up Clustering (Bottom-up clustering,BUC) [15], and softened similarity learning (SSL) [16]; 2) Unsupervised cross-domain methods. Multi-task mid-level feature alignment (MMFA) [17], Transferable joint attribute-identity deep learning (TJ-AIDL) [7], Triple adversarial learning and multi-view imaginative reasoning network (TAL-MIRN) [18] (Methods based on domain distribution alignment); Adaptive transfer network (ATNet) [9], similarity preserving generative adversarial network + local max pooling (Similarity preserving generative adversarial network + local max pooling, SPGAN+LMP) [19], heterogeneous − Hetero-homogeneous learning (HHL) [20] (GAN-based method); (Exemplar-invariance, camera-invariance and neighborhood-invariance,ECN) [12]example camera nearest neighbor, unsupervised domain adaptive person re-identification (Unsupervised domain adaptive person re-identification) -identification(UDAP) [21], Partbased convolutional baseline-progressive augmentation framework (PCB-PAST) with progressive enhancement framework, Self-similarity grouping (SSG) [10], based on progressive representation Enhanced self-training (Self-training with progressive representation enhancement, PREST) [22], soft iterative label clustering (SILC) [23]. All compared methodological results were obtained from the source papers. It can be seen from Tab.2 that the method in this paper is better than other methods.

V. LIMITATIONS AND CHALLENGES

Since network training in the target domain is statically affected by source gaps, it may affect the algorithm's ability to converge on the clustering of samples in the target domain. The unsupervised learning ability of the network can be further improved by fully considering the inter-domain distribution characteristics in the model and intervening dynamically in the training process in the target domain, which will serve as our subsequent research direction.

VI. CONCLUSION

Aiming at problems such as difficulty in obtaining data, expensive data labels, and weak model generalization ability in cross-domain person re-identification, a cross-domain person re-identification model based on normalized IBN-Net is proposed. By introducing the normalisation method to enhance the network's representational ability, important features in pedestrian images are effectively extracted. This improves the recognition accuracy and robustness, makes the training process more efficient, and brings significant performance improvement in the field of pedestrian re-identification. Triple optimization combines the IN layer

and BN layer and incorporates the SimAM attention mechanism to improve the model itself and better extract image features. Experimental results show that the pedestrian features extracted by this model are more discriminative and robust.The normalisation-based IBN-Net network model has achieved significant performance improvement in pedestrian re-identification. However, it still has some potential limitations, such as the generalisation ability to large-scale datasets which has not been fully validated, and high computational complexity. Future work can address these issues by further expanding the datasets to include more scenes, poses and occlusions to enhance the generalisation ability of the model, and by optimising the network structure and parameter settings to reduce the computational complexity, e.g., by introducing a lightweight network structure or adopting methods such as model pruning, in order to run more efficiently in practical applications.

REFERENCES

[1] Y. Yu, W. Zheng, L. Chao, H. Zhen, C. Jun, and H. Rui-Min, "A survey on multi-source person re-identification," *Acta Autom. Sinica*, vol. 46, no. 9, pp. 1869–1884, 2020.
[2] M. Ye, J. Shen, G. Lin, T. Xiang, L. Shao, and S. C. H. Hoi, "Deep learning for person re-identification: A survey and outlook," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 6, pp. 2872–2893, Jun. 2022.
[3] L. You-Jiao, Z. Li, Z. Jing, L. Jia-Feng, and Z. Hui, "A survey of person re-identification," *Acta Autom. Sinica*, vol. 44, no. 9, pp. 1554–1568, 2018.
[4] S. Bai, X. Bai, and Q. Tian, "Scalable person re-identification on supervised smoothed manifold," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2017, pp. 3356–3365.
[5] Z. Yun-Peng, W. Hong-Yuan, Z. Ji, C. Li, W. L. Yu, and G. Jia-Hui, "One-shot video-based person re-identification based on neighborhood center iteration strategy," *J. Softw.*, vol. 32, no. 12, pp. 4025–4035, 2021.
[6] L. Yi-Min, J. Jian-Guo, Q. Mei-Bin, L. Hao, and Z. H. Jie, "Video-based person re-identification method based on GAN and pose estimation," *Acta Autom. Sinica*, vol. 46, no. 3, pp. 576–584, 2020.
[7] J. Wang, X. Zhu, S. Gong, and W. Li, "Transferable joint attribute-identity deep learning for unsupervised person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2275–2284.
[8] D. Mekhazni, A. Bhuiyan, G. Ekladious, and E. Granger, "Unsupervised domain adaptation in the dissimilarity space for person re-identification," in *Proc. 16th Eur. Conf. Comput. Vis.*, 2020, pp. 159–174.
[9] J. Liu, Z.-J. Zha, D. Chen, R. Hong, and M. Wang, "Adaptive transfer network for cross-domain person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 7195–7204.
[10] Y. Fu, Y. Wei, G. Wang, Y. Zhou, H. Shi, U. Uiuc, and T. Huang, "Self-similarity grouping: A simple unsupervised cross domain adaptation approach for person re-identification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 6111–6120.
[11] Y. Zhai, S. Lu, Q. Ye, X. Shan, J. Chen, R. Ji, and Y. Tian, "AD-Cluster: Augmented discriminative clustering for domain adaptive person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 9018–9027.
[12] Z. Zhong, L. Zheng, Z. Luo, S. Li, and Y. Yang, "Invariance matters: Exemplar memory for domain adaptive person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 598–607.
[13] X. Pan, P. Luo, J. Shi, and X. Tang, "Two at once: Enhancing learning and generalization capacities via IBN-Net," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*. Cham, Switzerland: Springer, 2018, pp. 484–500.
[14] L. Yang, R. Y. Zhang, L. Li, and X. Xie, "SimAM: A simple, parameter-free attention module for convolutional neural networks," in *Proc. 38th Int. Conf. Mach. Learn.*, 2021, pp. 11863–11874.
[15] Y. Lin, X. Dong, L. Zheng, Y. Yan, and Y. Yang, "A bottom-up clustering approach to unsupervised person re-identification," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, no. 1, 2019, pp. 8738–8745.
[16] Y. Lin, L. Xie, Y. Wu, C. Yan, and Q. Tian, "Unsupervised person re-identification via softened similarity learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 3387–3396.
[17] S. Lin, H. Li, C. T. Li, and A. C. Kot, "Multi-task mid-level feature alignment network for unsupervised cross-dataset person re-identification," in *Proc. 29th Brit. Mach. Vis. Conf.*, 2018, pp. 1–9.
[18] H. Li, N. Dong, Z. Yu, D. Tao, and G. Qi, "Triple adversarial learning and multi-view imaginative reasoning for unsupervised domain adaptation person re-identification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 5, pp. 2814–2830, May 2022.
[19] W. Deng, L. Zheng, Q. Ye, G. Kang, Y. Yang, and J. Jiao, "Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 994–1003.
[20] Z. Zhong, L. Zheng, S. Li, and Y. Yang, "Generalizing a person retrieval model hetero-and homogeneously," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 172–188.
[21] L. Song, C. Wang, L. Zhang, B. Du, Q. Zhang, C. Huang, and X. Wang, "Unsupervised domain adaptive re-identification: Theory and practice," *Pattern Recognit.*, vol. 102, Jun. 2020, Art. no. 107173.
[22] H. Zhang, H. Cao, X. Yang, C. Deng, and D. Tao, "Self-training with progressive representation enhancement for unsupervised cross-domain person re-identification," *IEEE Trans. Image Process.*, vol. 30, pp. 5287–5298, 2021.
[23] J.-P. Ainam, K. Qin, J. W. Owusu, and G. Lu, "Unsupervised domain adaptation for person re-identification with iterative soft clustering," *Knowl.-Based Syst.*, vol. 212, Jan. 2021, Art. no. 106644.

XUEMEI BAI received the Ph.D. degree from Changchun University of Science and Technology, Changchun, China, in 2009. She is currently a Professor with the School of Electronic Information Engineering, Changchun University of Science and Technology. Her current research interests include intelligent information processing and pattern recognition.

AO WANG is currently pursuing the master's degree with the School of Electronic Information Engineering, Changchun University of Science and Technology. His research interest includes signal and information processing.

CHENJIE ZHANG received the master's degree from Changchun University of Science and Technology, in 2008. She is currently an Associate Professor with the School of Electronic Information Engineering, Changchun University of Science and Technology. Her current research interests include intelligent information processing and pattern recognition.

HANPING HU received the Ph.D. degree from Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, China, in 2015. He is currently a Laboratory Teacher with the School of Computer Science and Technology, Changchun University of Science and Technology. His current research interests include deep learning and pattern recognition.