## RESEARCH ARTICLE

# Unsupervised Cross-Domain Person Re-Identification Method Based on Attention Block and Refined Clustering

**YAN HUI**[ID]**, XI WU, XIUHUA HU, HUAN LIU, AND SHIJIE YOU**

State and Provincial Joint Engineering Laboratory of Advanced Network, Monitoring and Control, School of Computer Science and Engineering, Xi'an Technological University, Xi'an, Shaanxi 710021, China

Corresponding author: Yan Hui (yanxh_xatu@163.com)

**ABSTRACT** Most unsupervised cross-domain person re-identification methods based on clustering suffer from a lack of feature discrimination and clustering generates pseudo-labels noise, leading to a decrease in accuracy. To solve these problems, this paper proposes an unsupervised cross-domain person re-identification method based on attention block and refined clustering. Firstly, ResNet50 is selected as the backbone network, coordinate attention and triple attention are concatenated and embedded in ResNet50 to extract fine-grained features, perform feature aggregation, and mine fine-grained information. Secondly, a refined clustering strategy is proposed to achieve a coarse-to-fine clustering process by designing the measurement standards for clustering, determining its reliability, and eliminating noisy samples. Finally, the hybrid memory bank dynamically stores cluster centers and continues to update them with iterations, adapting to changes in clusters and performing invariant learning. The experimental results show that the new method designed in the paper improves the accuracy of rank-1 and mAP by 0.4% and 2.4%, respectively, on the target domain Market-1501 dataset, and improves the accuracy of rank-1 and mAP by 0.4% and 1.1%, respectively, on the target domain DukeMTMC-ReID dataset, compared with other typical methods.

**INDEX TERMS** Attention mechanism, cross-domain person re-identification, hybrid memory bank, refined clustering strategy, unsupervised methods.

## I. INTRODUCTION

Person re-identification, also known as pedestrian re-identification, aims at the process of re-identifying people captured by cameras deployed at different locations by whether they have the same identity information as the candidate. Person re-identification has strong practical significance, and it is widely used for smart security and social safety issues such as searching for lost children.

Person re-identification methods can be divided into study categories based on whether the image domain changes or not, including both same-domain and cross-domain types.

The associate editor coordinating the review of this manuscript and approving it for publication was Laura Celentano[ID].

Since the application of person re-identification techniques is infinite, the cross-domain approach is closer to practical applications and can adapt models trained in one environment to a new one, which is a valuable research problem. In recent years, person re-identification techniques have achieved good results in the same domain, but in the process of practical application, there are various types of image domain changes, such as different cameras, different weather, and different locations may have different image styles, so the model trained on the source domain dataset, when directly applied to the target domain dataset, the model is affected by cross-domain environmental factors and has obvious performance decline. Therefore, how to solve the problem of performance decline in cross-domain scenes due to image styles affected

Y. Hui *et al.*: Unsupervised Cross-Domain Person Re-Identification Method Based on Attention Block and Refined Clustering

IEEE *Access*

by environmental factors such as illumination, background, and viewpoint is the center of this study.

The unsupervised cross-domain strategy based on clustering generates pseudo-labels and uses the pseudo-labels for subsequent training. Compared with other unsupervised methods, the clustering-based approach has obvious advantages, as it does not require a lot of human and material resources to manually label the data, and there is no pre-modeling. However, it suffers from two problems, one is the pseudo-labels noise generated by clustering, and the other is the lack of feature discrimination.

The unsupervised person re-identification method based on clustering aims to learn unlabeled data on a large scale. At the very beginning of the study, the researchers did not consider the effect of cross-domain on the model, resulting in poor model extensibility. Lin *et al.* [3] proposed a bottom-up clustering (BUC) framework that iteratively trains networks with pseudo-labels. Ding *et al.* [4] proposed a clean and practical density-based clustering method by introducing a clustering effectiveness criterion. Subsequent analytical studies of the cross-domain problem introduced the clustering approach, but the new problem of pseudo-labels noise also emerged immediately. Fu *et al.* [5] proposed a self-similarity grouping (SSG) method, which automatically constructs multiple clusters from different views using the potential similarity of unlabeled samples. Li *et al.* [6] first assigned a multi-label vector to each image, and after several cycles of multi-label training, used clustering to assign a pseudo-label to each image. Wang *et al.* [7] used similarity calculation and circular consistency to ensure the quality of the pseudo-labels.

Although there were various methods for clustering, the models still suffered from noise, so in recent years researchers proposed a series of methods for overcoming pseudo-labels noise and generating reliable clusters. Yang *et al.* [8] eliminated outliers based on the clustering results to ensure that the model selects as many samples as possible during the training process. Yang *et al.* [9] improved the clustering algorithm by assigning pseudo-labels to noisy samples and proposed a dynamic and symmetric cross-entropy loss to resist the noisy labels generated by clustering. Zhai *et al.* [10] improved the discriminative power by adding clustering points to the target domain samples while expanding the data. Ge *et al.* [11] proposed a new framework for adaptive contrast learning, and the proposed adaptive approach gradually produces more reliable clusters to refine mixed memories and learning targets. Cho *et al.* [12] proposed a new framework for component-based pseudo-labels refinement to reduce label noise by employing complementary relationships between global and partial features. Chen *et al.* [13] proposed a person feature extraction method based on attention mechanism and contextual information fusion, and introduced a lightweight attention module in the ResNet50 backbone network to enhance salient human features and suppress irrelevant information by configuring a small number of network parameters. Although the above approaches have achieved better results using clus-

tering, two main issues still need to be addressed. First, there is the problem of domain gap, and the clusters generated by the extant methods are not reliable; Second, there is a large amount of noise in clusters, and some methods try to solve such problems by taking an extreme approach and choosing to simply discard the noisy samples and not use them for training anymore, however, the discarded samples may contain important person information, and a simple rough elimination will only damage the performance of the model.

For lack of feature discrimination, because there is a large gap between the illumination, background, and viewpoint in different environments, even the same person in different cameras, there will be differences in the extracted features. In the earliest studies, the main focus was on channel attention and spatial attention, with improvements on this basis. Jie *et al.* [14] proposed a "squeeze and excite" attention block that connects channel attention and spatial attention in tandem. Woo *et al.* [15] presented a new method to improve the representation capability of CNN networks, the convolutional bottleneck attention module. Zhang *et al.* [16] suggested a relationship-aware global attention module that obtains global structural information. In subsequent studies, researchers began to combine attribute information with models to implement the re-identification process. Tay *et al.* [17] proposed a unified learning framework that combined attribute information and attention. Chen *et al.* [18] presented a network architecture that integrated attention mechanisms with diversity regularization. Si *et al.* [19] suggested an attention network based on a hybrid memory bank, which can use global information to focus on identity features and train the model with attention. Then someone started to focus on location information to extract more fine-grained features. Zhong *et al.* [20] improved the robustness and accuracy of feature extraction by increasing the position and channel double attention mechanism. The above methods use attention mechanisms to capture channel, spatial, and attribute information, which relieves the lack of feature discrimination to some degree. However, for the loss of location information and cross dimensions, the above methods are rarely mentioned. Location information is crucial to the person re-identification algorithm, using it to select the space and cross dimensions to calculate attention weights to accurately predict the main areas of attention. Ignoring the location information and cross dimensions only damages the model accuracy and leads to a lack of feature discrimination.

In addition, the memory bank can be used to store information and improve the efficiency of model learning. In the earliest studies, the memory bank collected more feature information at the cost of storage space. Wang *et al.* [21] proposed a cross-batch memory mechanism that memorizes the embeddings of previous iterations. Zhong *et al.* [22] introduced a memory bank to store the features of the target domain and to contain the three invariant attributes. Part of the method improves the memory bank by introducing restriction constraints to ensure an invariant learning method. Luo *et al.* [23] improved the ordinary neighborhood invariant method

**IEEE** *Access*

Y. Hui *et al.*: Unsupervised Cross-Domain Person Re-Identification Method Based on Attention Block and Refined Clustering

by imposing constraints in a camera-aware approach. The above methods utilize the storage capacity of the memory bank for unsupervised learning, but they are not efficient due to excessive calculations.

For the above problem of person re-identification, the work taken in this paper is as follows:

(1) This paper proposes an unsupervised cross-domain person re-identification method based on attention blocks and refined clustering to solve the performance decline problem caused by cross-domain.

(2) In this paper, we design a refined clustering scheme to improve the quality of clusters, realize the coarse-to-fine clustering process, and solve the problem of pseudo-labels noise.

(3) In this paper, we introduce coordinate attention and triple attention to mine fine-grained features, and a hybrid memory bank to store cluster centers and continue to update them with iterations.

(4) The experimental results show that when tested on two publicly available datasets, Market-1501 and DukeMTMC-ReID, the accuracy of mAP is improved by 2.4% and 1.1%, respectively, which is significantly better than other existing methods.

## II. PROBLEM DESCRIPTION
Person re-identification techniques mainly include both supervised and unsupervised directions. The supervised approach is vulnerable to environmental factors, resulting in poor generalization ability. The unsupervised approach learns information about the labeled source domain data and the unlabeled target domain data, and the samples collected from the two datasets are different. Therefore, unsupervised person re-identification methods [1], [2] are increasingly important for practical applications.

The clustering-based approach is specific unsupervised learning, which has the obvious advantage of not requiring a lot of human and material resources to manually label the data and no pre-modeling. However, there are two problems, one is pseudo-labels noise, and the other is a lack of feature discrimination. Traditional unsupervised cross-domain person re-identification methods based on clustering [5], [6], firstly, use the backbone network ResNet50 for image feature extraction, rely on the extracted features, use DBSCAN or K-means clustering algorithms, and generate pseudo-labels, use the generated pseudo-labels to classify positive and negative instance samples, and then start unsupervised training using a triplet loss function. The framework is shown in Figure 1 and has two parts, i.e. backbone network to extract features and clustering to generate pseudo-labels, which are described in detail as follows.

The inputs to an unsupervised cross-domain person re-identification model are usually labeled source domain data and unlabeled target domain data. Since there are environmental differences between the source and target domains, the model is easily affected by illumination, background, and viewpoint. Even for the same person, there may be large differences in the physical features performed under non-overlapping cameras or different environmental conditions, while a different person may have similar body types or clothing, leading to a decline in feature discrimination. Therefore, how to better overcome the cross-domain problem is the key to this paper.

For the person re-identification method of clustering [2], [3], [4], [5], [6], the process of grouping data elements with more similarities into the same class by the clustering algorithm uses the DBSCAN algorithm to generate pseudo-labels for the unlabeled dataset. DBSCAN is a typical cluster algorithm based on density. The algorithm classifies data into three classes, cluster kernel points, boundary points, and outlier points. The detailed process is given the extracted features, calculates the Euclidean distance and calculates their Jaccard distance using K-nearest neighbors, and gets the Jaccard distance to generate pseudo-labels using DBSCAN. The above description is the main overview of the clustering

algorithm to solve the problem. The difference between various unsupervised cross-domain person re-identification methods based on clustering lies in the reliability of clustering and the deal with pseudo-labels noise. If the clustering is not measured and blindly deemed reliable, there will only be more and more noisy samples in the clusters; roughly eliminating the noise will cause the loss of information and destroy the accuracy of the model. Therefore how to perform reliability measures for clusters and extract information from noisy samples is one of the main issues in this paper.

For feature extraction, the traditional convolutional structure requires stacking many convolution layers to capture more local information. Although stacking more layers may improve the performance of these networks, the model will be prone to gradient vanishing and explosion due to the deeper network depth. Traditional person re-identification methods based on attention [16], [17], [18], [19], [20], use attention blocks to extract features of interest. The traditional approach achieves this by stacking a large number of attention blocks, leading to a heavy model calculation task, and the previous approach makes little mention of the absence of location information and cross dimensions. Therefore how to extract fine-grained features and solve the lack of feature discrimination is one of the main issues in this paper.

In view of the difficulties with existing techniques, this paper proposes an unsupervised cross-domain person re-identification method based on attention block and refined clustering to solve the above problems. We connect coordinate attention and triple attention in tandem, focus on location information, capture cross dimensions, and calculate attention weights. Coordinate attention focuses on location information and feature aggregation from two directions; triple attention is encoded using three branches, capturing cross dimensions and calculating attention weights. This paper also proposes a refining clustering algorithm to achieve a coarse-to-fine strategy to improve the reliability of clusters. In addition, because the cluster centers keep changing during the
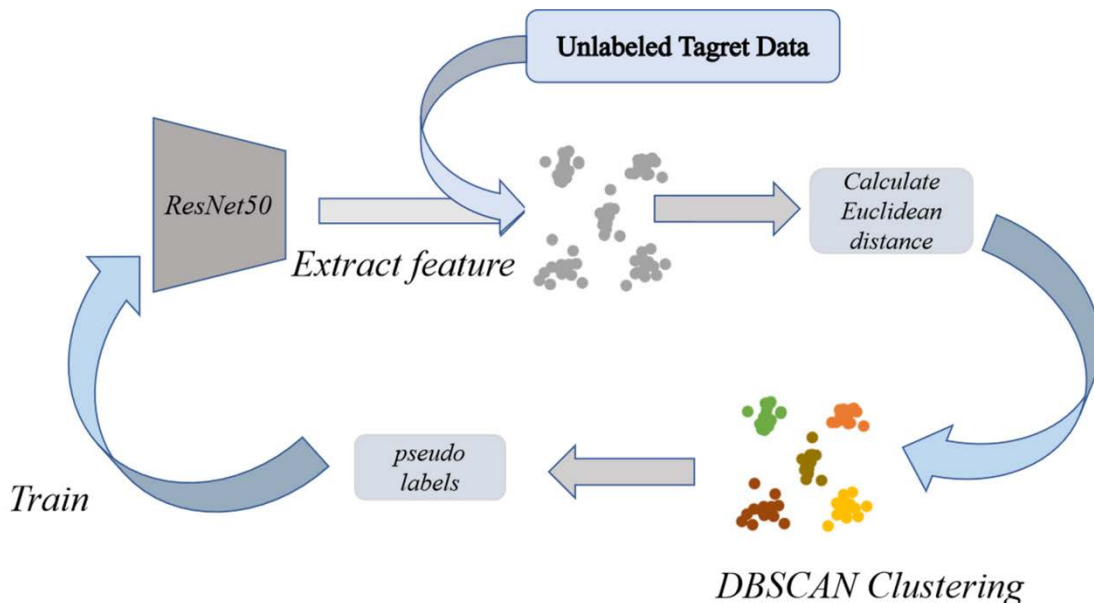
Y. Hui *et al.*: Unsupervised Cross-Domain Person Re-Identification Method Based on Attention Block and Refined Clustering

**IEEE** *Access*

training process, this paper uses a hybrid memory bank to store the cluster centers for invariant learning.

## III. METHOD

For the cross-domain problem, pseudo-labels noise, and lack of feature discrimination, this paper proposes an unsupervised cross-domain person re-identification method based on attention block and refined clustering, and the framework figure of this method is shown in Figure 2. This framework has four parts, which are the backbone model, attention block, clustering algorithm, and memory bank. ResNet50 is used as the backbone network, the attention block is a tandem of coordinate attention and triple attention embedded before layer1 and after layer4 of ResNet50 to extract fine-grained features. This paper proposes a refined strategy using the DBSCAN algorithm to design a reliability measure standard to achieve a coarse-to-fine clustering process. The hybrid memory bank dynamically stores cluster centers and continues to update them with iterations, adapting to changes in clusters and performing invariant learning.

The framework is input with labeled source and unlabeled target domains, and ResNet50 is selected as the backbone model to embed the attention block before layer1 and after layer4 of ResNet50. The attention block in this paper is a tandem of coordinate attention and triple attention, with coordinate attention in the first and triple attention in the second. Coordinate attention focuses on location information and triple attention capturing cross dimensions, through improvements to the backbone network to overcome the lack of feature discrimination. After that, the extracted features are clustered using DBSCAN to achieve a coarse-to-fine clustering process by calculating independence scores and performing reliability measures to overcome the problem of

pseudo-labels noise. Meanwhile, the hybrid memory bank performs a feature store, which is updated continuously to adapt to the changes in clusters. Finally, this paper uses the pseudo-labels generated by refined clustering to calculate the Euclidean distance between samples, determine positive and negative samples, perform metric learning using triplet loss, and perform classification using cross-entropy loss.

### A. ATTENTION BLOCK

ResNet50 is used as the backbone network to extract image features. It contains 1 convolution layer and 4 residual modules, each residual module contains multiple convolution layers, batch normalization layers, and activation functions. Because the influence of illumination, background, and viewpoint causes the features extracted by ResNet50 to lack discrimination and the model accuracy to be poor, attention block are introduced in this paper to extract more fine-grained features.

In this paper, the structure of the backbone network is fine-tuned by adding the attention block before layer1 and after layer4. Before the deep convolution of the model begins, the attention block is used to focus on the location information of the samples for initial feature extraction for subsequent convolution training. After that, feature mining is performed using an attention block to extract more fine-grained features.

To address the problem of lack of feature discrimination, attention blocks are added to the backbone network to extract more fine-grained features. In this paper, coordinate attention and triple attention are concatenated, where coordinate attention is used to reduce the loss of location information due to global pooling, and triple attention captures the cross dimensions and calculates the attention weights.
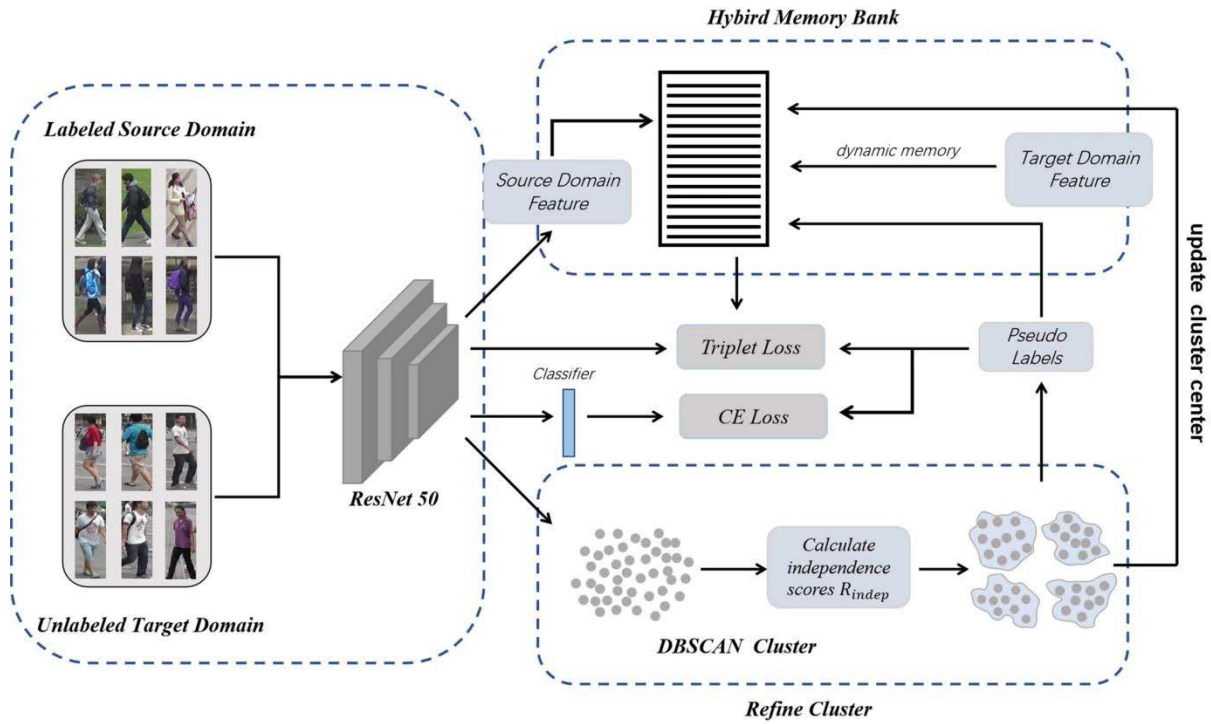
IEEE *Access*

Y. Hui *et al.*: Unsupervised Cross-Domain Person Re-Identification Method Based on Attention Block and Refined Clustering



**FIGURE 2.** Framework of unsupervised cross-domain person re-identification method based on attention block and refined clustering.

## 1) COORDINATE ATTENTION

The previous approach mitigates the lack of feature discrimination to some extent by focusing on channel, spatial, and attribute information. However, there are a few existing methods mentioned for the loss of location information and cross dimensions. In this paper, coordinate attention is used to reduce the loss of location information due to global pooling. The feature structure is shown in Figure 4, which includes two parts: coordinate information embedding and coordinate information generation.

For coordinate attention, first given the input $X$, each channel is encoded along the horizontal and vertical coordinates using pooling kernels of size $(H, 1)$ or $(1, W)$, respectively. Thus, the output of the $c$-th channel with height $h$ can be expressed as:

$$Z_c^h (h) = \frac{1}{W} \sum_{0 \le j < W} x_c (h, i) \tag{1}$$

where $Z_c^h (h)$ denotes the output of the $c$-th channel at height $h$. Similarly, the output of the $c$-th channel with width $w$ can be expressed as:

$$Z_c^w (w) = \frac{1}{H} \sum_{0 \le j < H} x_c (j, w) \tag{2}$$

where $Z_c^w (w)$ denotes the output of the $c$-th channel at width $w$. Then the feature maps along both horizontal and vertical directions are concatenated together, after which they are fed into a $1 \times 1$ convolution to generate a middle feature map $m$,

which is represented as:

$$m = \delta \left( F \left( \left[ Z^h, Z^w \right] \right) \right) \tag{3}$$

where $[,]$ denotes the tandem operation, $\delta$ is a nonlinear activation function, and $F$ is a $1 \times 1$ convolution variation function. Then the spatial dimension is divided into two independent tensors, and the two independent tensors are varied using two other $1 \times 1$ convolutions to obtain two tensors with the same number of channels as the input $X$.

$$g^h = \text{sigmoid} \left( F_h \left( m^h \right) \right) \tag{4}$$

$$g^w = \text{sigmoid} \left( F_w \left( m^w \right) \right) \tag{5}$$

where sigmoid () denotes the sigmoid activation function.

Finally, the output of the coordinate attention block $Y$ is represented as:

$$Y (i, j) = x (i, j) \times g^h \times g^w \tag{6}$$

## 2) TRIPLE ATTENTION

For the absence of cross dimensions, this paper introduces triple attention to capture the cross dimensions and calculate the attention weights, the structure of which is shown in Figure 5. The triple attention consists of three branches, two of which are responsible for capturing the cross-dimensional interactions between the channel dimension and the spatial dimension, and the last branch is used to establish spatial attention.

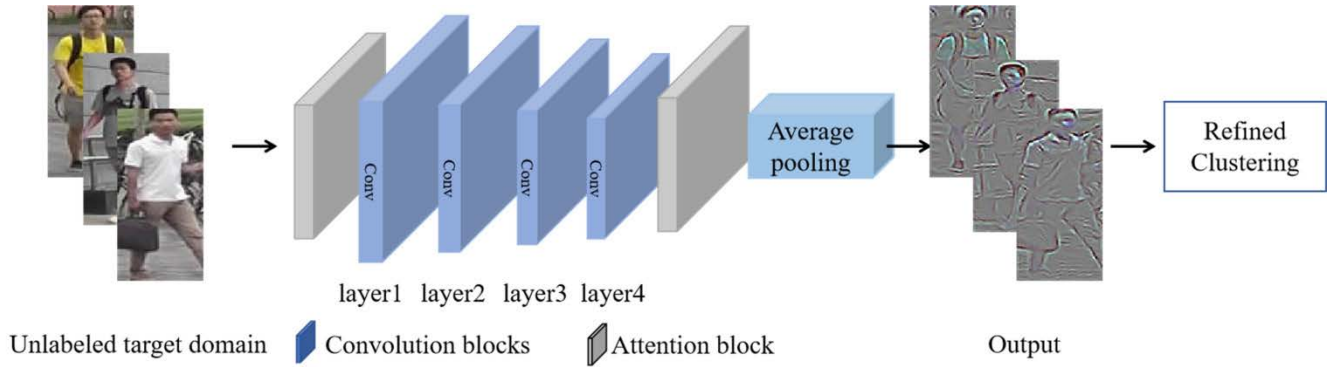For triple attention, given the input $Y$, pass it to each branch. The first branch creates an interaction between the

Y. Hui *et al.*: Unsupervised Cross-Domain Person Re-Identification Method Based on Attention Block and Refined Clustering

**IEEE** *Access*



**FIGURE 3.** Diagram of backbone network details.
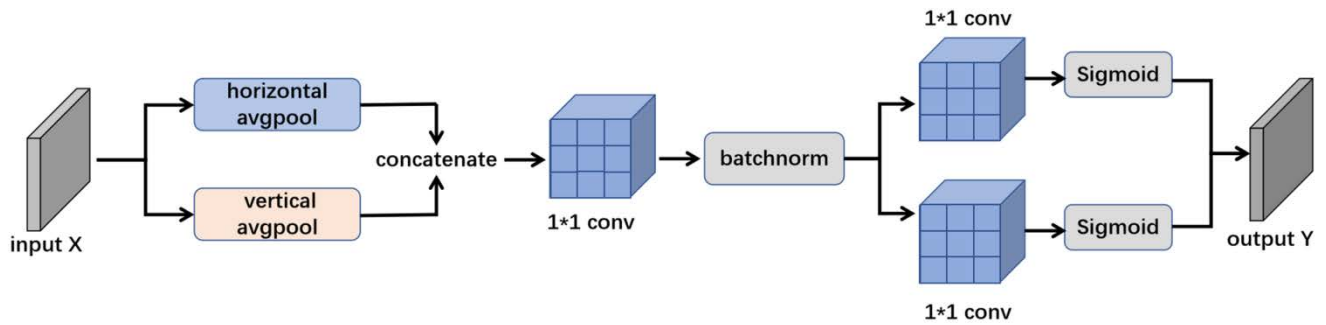


**FIGURE 4.** Chart of coordinate attention structure.



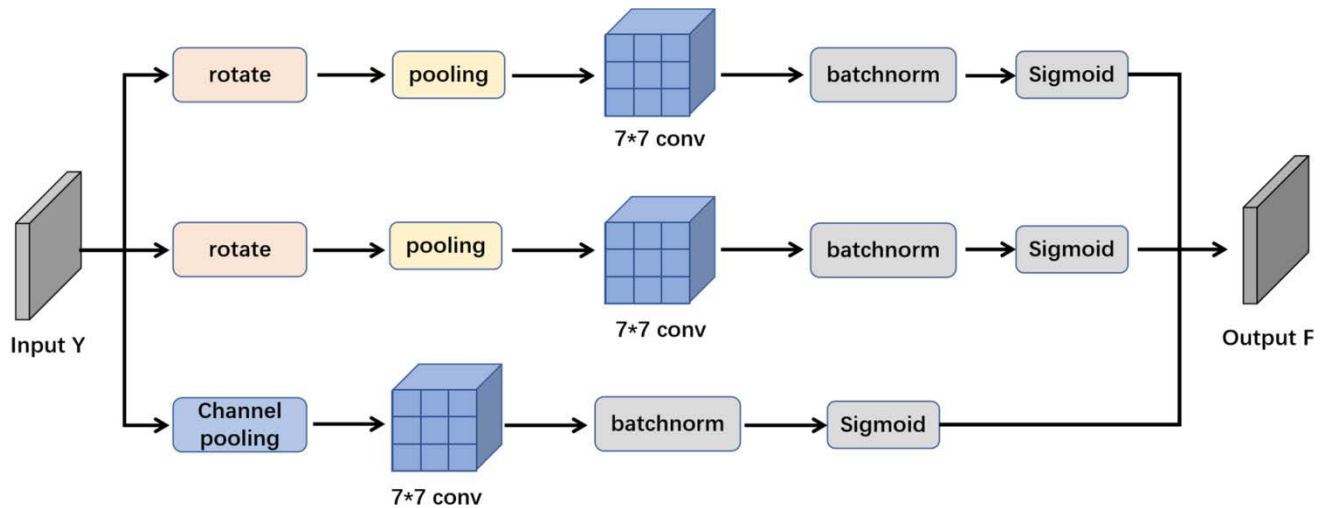**FIGURE 5.** Chart of triple attention structure.

height and channel dimensions. The input $Y$ is rotated 90° counterclockwise along the $H$-axis, and this rotation tensor is represented as $a_1$, which is subsequently passed through a pooling layer and represented as $a_1^*$. and then $a_1^*$ is fed to the $7 \times 7$ convolution and batch normalization layers. Finally, the resulting attention weights are produced by passing the tensor through the sigmoid function.

For the second branch, the input $Y$ is rotated 90° counterclockwise along the $W$-axis, and this rotation tensor is represented as $a_2$, which is subsequently passed through a pooling layer and represented as $a_2^*$. and then $a_2^*$ is fed to the $7 \times 7$ convolution and batch normalization layers. Finally, the attention weights are produced by the sigmoid function.

IEEE Access

Y. Hui *et al.*: Unsupervised Cross-Domain Person Re-Identification Method Based on Attention Block and Refined Clustering

In the third branch, the input $Y$ does not require any rotation operation and is directly channel pooling with the tensor represented as $a_3$, which is then fed into a $7 \times 7$ convolution kernel, a batch normalization layer. Finally, the sigmoid function is used to generate the attention weights.

The equation formula expression for the whole process as:

$$F(i,j) = \frac{1}{3}\left(a_1^* w_1 + a_2^* w_1 + a_3 w_3\right) \tag{7}$$

where $w_1$, $w_2$ and $w_3$ denote the attention weights of three different branches, respectively.

### B. REFINED CLUSTERING ALGORITHM

Previous unsupervised cross-domain person re-identification methods based on clustering inevitably have outliers in the process of generating pseudo-labels. They used simple elimination to deal with the noise problem, but this only destroys the accuracy of the model. For the problem of pseudo-labels noise generated by clustering, this paper proposes a refined clustering strategy. In this paper, a reliability measure is designed to determine the reliability of clustering instances, release the interference of noisy samples, and realize the clustering process from coarse to fine.

The specific process is as follows: firstly, using the feature $F(i,j)$ extracted by the backbone network, Jaccard distance is calculated, and preliminary clustering is performed. In clustering, different clustering radii are set to produce three clusters with the same cluster center but different radii. To get its independence score for it as IoU (Intersection over Union), a reliability measure is performed to determine the reliability of the clustering instances, and then pseudo labels are generated. Before the start of the next iteration, the model retains only reliable clusters, and the remaining unreliable clusters and samples are considered outliers. Outliers will continue to cluster in subsequent iterations until reliable clusters can be produced, thus achieving a coarse-to-fine clustering process. In this paper, we use a refined clustering method to obtain pseudo-labels for reliable clustering to overcome the noise problem and improve model accuracy without losing any information.

The measure of cluster independence is defined by the IoU score.

$$R_{indep}(f_i^t) = \frac{\left|I(f_i^t) \cap I_{loose}(f_i^t) \cap I_{tight}(f_i^t)\right|}{\left|I(f_i^t) \cup I_{loose}(f_i^t) \cup I_{tight}(f_i^t)\right|} \tag{8}$$

where $I(f_i^t)$ denotes the samples of $f_i^t$ within the same cluster, $I_{loose}(f_i^t)$ is the set of clusters containing $f_i^t$ when the clustering condition becomes loose, and $I_{tight}(f_i^t)$ is the set of clusters containing $f_i^t$ when the clustering condition becomes tight. A larger $R_{indep}(f_i^t)$ indicates that the cluster is more $f_i^t$ independent, where $R_{indep}(f_i^t)$ has a value between 0 and 1.

Before the start of the next iteration, the clusters are filtered according to the reliability measure, and the clusters with independence scores between 0 and 1 are considered reliable clusters, and the rest are unreliable clusters. The model retains only reliable clusters, and all other unreliable clusters and

samples are considered outliers. In subsequent iterations, the outliers will continue to cluster until reliable clusters are produced. As the number of iterations increases, unreliable clusters are gradually filtered based on the reliability measure standard to achieve a coarse-to-fine clustering process. In addition, the center points of the reliable clusters are stored in the hybrid memory bank for subsequent training.

### C. HYBRID MEMORY BANK

To improve the generalization ability of the network over the target domain dataset, invariant learning of the network by estimating the similarity between target images is proposed.

For source domain data, because of having real classes, this paper proposes to store them in classes. The source domain features within the current small batch are counted as mean values according to the classes and then accumulated as momentum to the corresponding class mass centers in the hybrid memory bank.

For the target domain data, this paper proposes to store the features all in sample instances, which allows the target domain samples to be continuously updated in the hybrid memory bank even when the clusters and outliers are constantly changing. The target domain features within the current small batch are accumulated to the corresponding instance features of the hybrid memory bank according to the index of the instances.

The specific procedure, which starts with the initialization of the memory bank, treats each cluster as a class and uses the average features of each cluster to initialize the class-level store $K[i]$. The specific update is expressed as follows.

$$K[i] = \frac{1}{|I_k|} \sum_{v_i \in I_k} v_i \tag{9}$$

where $I_k$ denotes the reliable cluster containing feature instance $x_i$, $K[i]$ denotes the initialized class-level store, $||$ denotes the number of features in this cluster, and $v_i$ denotes the target domain feature instance.

Then a module update is performed to construct a memory bank to store the cluster center points based on all instances in the $k$th cluster, which is continuously updated at each iteration. After each clustering, it is updated to feed the centers of reliable clusters into the module. In the hybrid memory bank, the cluster prime is obtained by averaging the features of the same cluster-ID. And the outlier instance features are extracted directly from the hybrid memory bank for the remaining instance features.

$$K[i] \leftarrow \beta K[i] + (1-\beta)v_i \tag{10}$$

where $\beta \in [0,1]$ is the update rate.

The hybrid memory bank performs feature storage, allowing the target domain samples to be continuously updated even when the clusters and non-clustered outliers are constantly changing, storing the clustering centers to accommodate the changes in clusters. In addition, as the source and target domain, sample features are updated from the hybrid

Y. Hui *et al.*: Unsupervised Cross-Domain Person Re-Identification Method Based on Attention Block and Refined Clustering

IEEE *Access*

memory bank, more reliable clusters can be gradually filtered out to further improve feature learning.

### D. ALOGORITHM IMPLEMENTATION

In the whole train process, the network is first pre-trained by inputting labeled source domain images, followed by inputting unlabeled target domain images into the model for training. Attention blocks are embedded in the backbone network to extract fine-grained features. The refined clustering algorithm performs feature clustering to achieve a coarse-to-fine clustering process based on a reliability measure. Memory bank to store cluster center points and adapt to cluster changes.

Input: labeled source domain images, unlabeled target domain images

Output: mAP, Rank accuracy

The specific implementation process is as follows.

*Step 1:* Input labeled source domain images and pre-train the network model.

*Step 2:* Input the unlabeled target domain image and continue training on the already trained model.

*Step 3:* In the feature extraction stage, the location attention and triple attention are concatenated to focus on the location information and capture the cross dimensions to calculate the attention weights to obtain feature $Y(i,j)$ based on Eq. (6). then the fine-grained feature $F(i,j)$ will be obtained based on Eq. (7) using triple attention.

*Step 4:* The extracted features $F(i,j)$ are clustered by using the refined clustering method for feature clustering, setting three different cluster radii, using the IoU score for reliability measure, and calculating the independence score $R_{indep}(f_i^t)$ based on Eq. (8).

*Step 5:* Clusters with independence scores between 0 and 1 are considered reliable clusters, and unreliable clusters are re-clustered with noisy instances in the next iteration.

*Step 6:* Initialize the hybrid memory bank according to Eq. (9), feed the cluster centers of reliable clustering into the module, adapt to cluster changes, and update the memory bank $K[i]$ according to Eq. (10).

*Step 7:* Train the model until the end of the iteration.

## IV. EXPERIMENTS

To analyze the performance of the proposed method on two publicly available person re-identification datasets, DukeMTMC-ReID [24] and Market-1501 [25], the entire network is built using the Pytorch framework with an input image size of $256 \times 128$ pixels. Random data enhancement is applied to each image before it is fed into the network, including random flips, crops, and erasures. The source domain is pre-trained for 50 epochs. $batchsize_1$ is set to 128, the initial learning rate $\alpha_1$ is set to 0.0003, the target domain is trained for 70 epochs, the initial learning rate $\alpha_2$ set to 0.00035. $batchsize_2$ set to 64, the original clustering radius eps is 0.6, the sparse clustering radius eps_loose is 0.62, the compact clustering radius eps_tight is 0.58, noted as $\Delta d=0.02$, the hyperparameter of Jaccard distance $k_1$ is set to 30, the $k_2$ is

**TABLE 1.** Analysis table of the selected positions of the attention block.

| Selected location | Target Domain：Market-1501 | | | |
| --- | --- | --- | --- | --- |
| | mAP | Rank-1 | Rank-5 | Rank-10 |
| Layer1+layer2 | 69.0 | 83.9 | 93.6 | 95.6 |
| Layer1+layer3 | 70.1 | 86.0 | 93.9 | 95.4 |
| Layer2+layer3 | 70.1 | 85.5 | 93.8 | 95.3 |
| Layer2+layer4 | 70.5 | 85.9 | 93.1 | 95.4 |
| Layer3+layer4 | 70.1 | 85.8 | 93.7 | 95.3 |
| Layer1+layer4 | 70.7 | 87.1 | 94.0 | 96.4 |

set to 6, and the update rate of the memory bank $\beta$ is set to 0.2.

To evaluate the retrieval performance of the person re-identification algorithm for query images, the mAP and Rank [26] accuracy metrics were used to measure algorithm retrieval results, with higher values indicating higher accuracy of the person re-identification model.

### A. ATTENTION BLOCK ANALYSIS

In this paper, coordinate attention is introduced to extract more fine-grained features. Coordinate attention is added before layer1 and after layer4 of ResNet50. As shown in Table 1, the source domain is DukeMTMC-ReID, and the target domain is Market-1501.

The attention blocks are set at different locations in the backbone network to justify the locations selected in this paper. Layer-n indicates the location where the attention blocks are embedded. In this paper, the attention block is embedded before layer1 and after layer4 of ResNet50. Before the deep convolution of the model begins, the location information is attended to using the coordinate attention block, the triple attention captures the cross dimensions, and the attention weights are calculated for feature extraction to facilitate subsequent convolution training. After the model has extracted the features, feature mining is then performed using attention blocks to extract more fine-grained features.

According to the detailed analysis of the table, "Layer1 + Layer4" has the best effect, while "Layer1 + Layer2" has the worst effect. The reason is that the features extracted by Layer1 and Layer2 experienced a small number of convolution boxes, and the features were not all extracted, so the display performance was poor. "Layer2+ Layer4" is slightly inferior to the optimal solution, but it also shows strong performance, but the loss of some of the original features is the reason for its poor performance. "Layer1 + Layer4" means that initial features are collected in the Layer1 layer and fine-grained features are extracted in the Layer4 layer, so the performance is better.

To visualize the effectiveness of the attention block, the pedestrian images taken in the Market-1501 dataset under complex scenarios such as camera view changes, different

**IEEE** *Access*

Y. Hui *et al.*: Unsupervised Cross-Domain Person Re-Identification Method Based on Attention Block and Refined Clustering

**FIGURE 6.** Diagram of attention visualization.
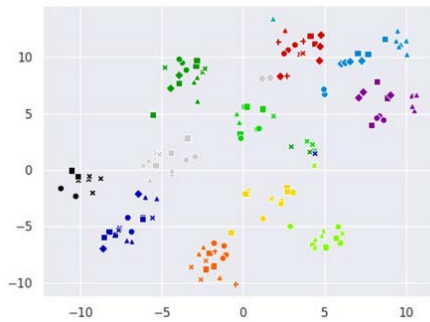


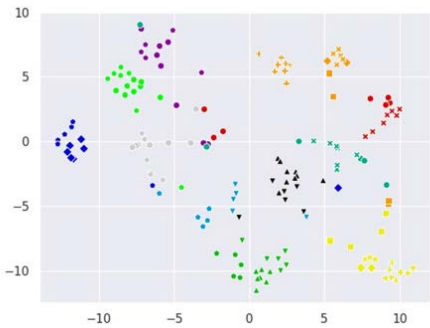**FIGURE 7.** Visualization of Market-1501 dataset in the target domain.



**FIGURE 8.** Visualization of DukeMTMC-ReID dataset in the target domain.

pedestrians dressed similarly, illumination changes, and the presence of occlusion were selected to visualize the areas of attention block, and the results are shown in Figure 6. The dark red colored areas are the main areas of attention in the network.

### B. CLUSTER VISUALIZATION ANALYSIS

As shown in Figure 7 and Figure 8, this paper proposes a refined clustering strategy to determine whether the clusters are reliable by designing a reliability measure to perform the selection of clusters, and the clusters with independence scores between 0 and 1 are regarded as reliable clusters, and the rest are regarded as unreliable clusters. The outliers will continue to cluster in the subsequent process until reliable clusters are produced. This coarse-to-fine clustering method further improves the model performance and the clustering

**TABLE 2.** Experimental results of different modules under the target domain Market-1501.

| Methods | DukeMTMC-ReID→Market-1501 | | | |
|---|---|---|---|---|
| | mAP | Rank-1 | Rank-5 | Rank-10 |
| Baseline+Cluster | 48.0 | 72.2 | 83.6 | 87.1 |
| Baseline+AT+Cluster | 50.0 | 72.7 | 85.5 | 88.6 |
| Baseline+Refined-Cluster | 68.2 | 86.0 | 93.4 | 95.4 |
| Baseline+AT+Refined-Cluster | 70.7 | 87.1 | 94.0 | 96.4 |

**TABLE 3.** Experimental results of different modules under the target domain DukeMTMC-ReID.

| Methods | Market-1501→DukeMTMC-ReID | | | |
|---|---|---|---|---|
| | mAP | Rank-1 | Rank-5 | Rank-10 |
| Baseline+Cluster | 46.4 | 66.4 | 77.7 | 81.7 |
| Baseline+AT+Cluster | 47.7 | 67.0 | 78.9 | 82.3 |
| Baseline+Refine-Cluster | 60.3 | 76.9 | 85.9 | 88.5 |
| Baseline+AT+Refine-Cluster | 61.5 | 77.6 | 87.0 | 90.2 |

effect is significantly enhanced. The following is a visual demonstration of its clustering effect.

### C. METHOD ANALYSIS IN THIS PAPER

The ablation experiments in this paper were conducted on two publicly available datasets, Market-1501 and DukeMTMC-ReID. To explore the effects of different modules on the experimental results, the Rank and mAP obtained from each group of experiments are shown in Table 2 and Table 3. Where Baseline denotes the backbone network of ResNet50, AT denotes attention block, Cluster denotes the DBSCAN clustering method, and Refined-Cluster denotes the method of refined clustering. "Baseline+Cluster" means that ResNet50 and DBSCAN are used for identification, "Baseline+AT+Cluster" denotes the backbone network fed with attention block after training, and "Baseline+Refined-Cluster" indicates that the refined clustering method proposed in this paper is used for training, "Baseline+AT+Refined-Cluster" indicates that the attention block and refined clustering methods.

From Table 2, we can learn that the mAP of "Baseline+Cluster" is only 48.0% in the source domain of DukeMTMC-ReID and the target domain of Market-1501, and the reason for the low accuracy of the model at this time is the pseudo-labels noise generated by clustering and the lack of feature discrimination. In contrast, when the coordinate attention and triple attention are embedded in the model, the mAP is 50.0%, which shows that the proposed features have slightly less impact on the model when the clustering method is not changed. When a refined clustering method is applied to the model, the mAP increases substantially to 68.2%. Thus, the coordinate attention overcomes the lack

Y. Hui et al.: Unsupervised Cross-Domain Person Re-Identification Method Based on Attention Block and Refined Clustering

IEEE Access

**TABLE 4.** Performance comparison with related methods on the Market-1501 dataset.

| Methods | Market-1501 | | | | |
|---|---|---|---|---|---|
| | Source | mAP | Rank-1 | Rank-5 | Rank-10 |
| BUC[3] | None | 38.3 | 66.2 | 79.6 | 84.5 |
| DBC[4] | None | 41.3 | 69.2 | 83.0 | 87.8 |
| SSL[28] | None | 37.8 | 71.7 | 83.8 | 87.4 |
| HCT[27] | None | 56.4 | 80.0 | 91.6 | 95.2 |
| MMCL[7] | Duke | 60.4 | 84.4 | 92.8 | 95.0 |
| SSG[5] | Duke | 58.3 | 80.0 | 90.0 | 92.4 |
| MLC[6] | Duke | 65.9 | 85.6 | 93.9 | 96.0 |
| ADTC[31] | Duke | 59.7 | 79.3 | 90.8 | 94.1 |
| AD-Cluster[10] | Duke | 68.3[2] | 86.7[2] | 94.4[2] | 96.5[2] |
| D-MMD[29] | Duke | 48.8 | 70.6 | 87.0 | 91.5 |
| TJ-AIDL[30] | Duke | 26.5 | 58.2 | 74.8 | 81.1 |
| HHL[32] | Duke | 31.4 | 62.2 | 78.8 | 84.0 |
| SCCT[33] | Duke | 67.9[3] | 86.6[3] | 94.5[1] | 96.9[1] |
| Ours | Duke | 70.7[1] | 87.1[1] | 94.0[3] | 96.4[3] |

**TABLE 5.** Performance comparison with related methods on the DukeMTMC-ReID dataset.

| Methods | DukeMTMC-ReID | | | | |
|---|---|---|---|---|---|
| | Source | mAP | Rank-1 | Rank-5 | Rank-10 |
| BUC[3] | None | 27.5 | 47.4 | 62.6 | 68.4 |
| DBC[4] | None | 30.0 | 51.5 | 64.6 | 70.1 |
| SSL[28] | None | 28.6 | 52.5 | 63.5 | 68.9 |
| HCT[27] | None | 50.7 | 69.6 | 83.4 | 87.43 |
| MMCL[7] | Market | 16.2 | 43.6 | 54.3 | 58.9 |
| SSG[5] | Market | 53.4 | 73.0 | 80.6 | 83.2 |
| MLC[6] | Market | 55.0[3] | 74.1[3] | 83.8 | 86.3 |
| ADTC[31] | Market | 52.5 | 71.9 | 84.1[3] | 87.5[3] |
| AD-Cluster[10] | Market | 54.1 | 72.6 | 82.5 | 85.5 |
| D-MMD[29] | Market | 46.0 | 63.5 | 78.8 | 83.9 |
| TJ-AIDL[30] | Market | 23.0 | 44.3 | 59.6 | 65.0 |
| HHL[32] | Market | 27.2 | 46.9 | 61.0 | 66.7 |
| SCCT[33] | Market | 60.4[2] | 76.0[2] | 85.0[2] | 88.9[2] |
| Ours | Market | 61.5[1] | 77.6[1] | 87.0[1] | 90.2[1] |

of feature discrimination, the triple attention captures the cross dimensions and calculates the attention weights, and the refined clustering approach overcomes the pseudo-labels noise. If both attention block and refined clustering methods are added to the model, with the former mining fine-grained features and the latter achieving a coarse-to-fine clustering process, the accuracy of the model is optimized with an mAP of 70.7%.

From Table 3, we can learn that the mAP of "baseline+Cluster" is only 46.4% in the source domain of Market-1501 and the target domain of DukeMTMC-Reid. In comparison, when attention blocks are embedded in the model, the mAP reaches 47.7%; when the method of refined clustering is applied to the model, the mAP increases substantially to 60.3%. If the methods of attention block and refined clustering are added to the model at the same time, the model accuracy achieves the best result with an mAP of 61.5%. In this paper, attention block is used to solve the problem of lack of feature discrimination, while the problem of pseudo-labels noise is solved by a coarse-to-fine clustering method.

## D. COMPARISON WITH EXISTING METHODS
The performance of the proposed algorithm is compared with typical unsupervised cross-domain pedestrian re-identification methods on the Market-1501 public dataset, where 12 comparison methods including BUC[3], DBC[4], SSL[28], MLC[6], MMCL[7], AD-Cluster[10], HCT[27], SSL[28], D-MMD[29], TJ-AIDL[30], ADTC[31], HHL[32], and SCCT[33] are analyzed based on the results of evaluation metrics mAP and Rank to verify the effectiveness of the proposed pedestrian re-identification algorithm.

As can be seen from Table 4, when the target domain is Market-1501, the model mAP of this paper is 70.7% and Rank-1 is 87.1%, which is an improvement compared with other methods. Specifically, compared with pure cluster methods such as BUC, DBC, and HCT, the performance of the method proposed in this paper is significantly better than those pre-trained by source domains. Compared with unsupervised cross-domain methods such as MMCL, MLC, and D-MMD, the method proposed in this paper uses an unsupervised cross-domain method based on clustering and uses refined clustering for training to achieve a coarse-to-fine clustering process that focuses on the information of noisy samples and does not discard noisy samples. Compared with SSG, a clustering method, coordinate attention is used to solve the problem of lack of feature discrimination, and coarse and refined clustering methods are used to solve the problem of pseudo-labels noise. Compared with the based on GAN methods PTGAN and TJ-AIDL, the method in this paper achieves good performance in both mAP accuracy and Rank, which shows that the proposed method in this paper far exceeds the based on GAN methods.

As can be seen from Table 5, when the target domain is DukeMTMC-ReID, the model in this paper has an mAP of 61.5% and a Rank-1 of 77.6%, ranking first in each index compared to other methods. Specifically, compared with the pure clustering HCT, DBC, and HCT methods, the performance of the proposed method in this paper is still better than it, with a 10.8% improvement in mAP. Compared with unsupervised cross-domain methods such as MLC, ADTC, and AD-Cluster, this paper shows good performance by combining the refined clustering method with attention block, with a 6.5% improvement in mAP. Compared with the SSG

IEEE Access

Y. Hui *et al.*: Unsupervised Cross-Domain Person Re-Identification Method Based on Attention Block and Refined Clustering

unsupervised cross-domain clustering method, the attention block and coarse-to-fine clustering methods not only solve the problem of lack of feature discrimination but also solve the problem of pseudo-labels noise. Compared with the based GAN methods PTGAN and TJ-AIDL, which shows that the proposed method in this paper far exceeds the based GAN methods. Compared with the HHL generative adversarial network-based method, the method in this paper is 34.3% higher in mAP accuracy.

## V. CONCLUSION

In this paper, we propose an unsupervised cross-domain person re-identification method based on attention block and refined clustering. The method is based on the ResNet50 network, which connects coordinate attention and triple attention in tandem to extract more discriminative features. This paper also proposes a refined clustering algorithm to achieve a coarse-to-fine strategy to improve the reliability of clusters. In addition, because the cluster centers keep changing during the training process, this paper uses a hybrid memory bank to store the cluster centers for invariant learning. The proposed method solves the problems of lack of feature discrimination and clustering generated by pseudo-labels noise. Full experiments are conducted on Market-1501 and DukeMTMC-ReID public datasets and compared with related algorithms, and the results show that the proposed method has the advantages of high accuracy and robustness in the unsupervised cross-domain scene. In future work, we will further study how to perform fast image matches and further improve the generalization of the re-identification model.

## REFERENCES

[1] Q. Leng, M. Ye, and Q. Tian, "A survey of open-world person re-identification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 4, pp. 1092–1108, Apr. 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] Y. Lin, X. Dong, L. Zheng, Y. Yan, and Y. Yang, "A bottom-up clustering approach to unsupervised person re-identification," in *Proc. 33rd AAAI Conf. Artif. Intell.*, Honolulu, HI, USA, Jan. 2019, pp. 8738–8745.

[4] G. Ding, S. Khan, Z. Tang, J. Zhang, and F. Porikli, "Towards better validity: Dispersion based clustering for unsupervised person re-identification," in *Proc. BMVC*, Cardiff, U.K., Sep. 2019, pp. 8–18.

[5] 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.

[6] Q. Li, X. Peng, Y. Qiao, and Q. Hao, "Unsupervised person re-identification with multi-label learning guided self-paced clustering," *Pattern Recognit.*, vol. 125, May 2022, Art. no. 108521.

[7] D. Wang and S. Zhang, "Unsupervised person re-identification via multi-label classification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Seattle, WA, USA, Jun. 2020, pp. 10978–10987.

[8] F. Yang, K. Li, Z. Zhong, Z. Luo, X. Sun, H. Cheng, X. Guo, F. Huang, R. Ji, and S. Li, "Asymmetric co-teaching for unsupervised cross domain person re-identification," in *Proc. 34th AAAI Conf. Artif. Intell.*, New York, NY, USA, Feb. 2020, pp. 12597–12604.

[9] F. Yang, Z. Zhong, Z. Luo, Y. Cai, Y. Lin, S. Li, and N. Sebe, "Joint noise-tolerant learning and meta camera shift adaptation for unsupervised person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 4855–4864.

[10] 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)*, Seattle, WA, USA, Jun. 2020, pp. 9021–9030.

[11] Y. Ge, D. Chen, F. Zhu, R. Zhao, and H. Li, "Self-paced contrastive learning with hybrid memory for domain adaptive object re-ID," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2022, pp. 11309–11321.

[12] Y. Cho, W. J. Kim, S. Hong, and S.-E. Yoon, "Part-based pseudo label refinement for unsupervised person re-identification," 2022, *arXiv:2203.14675*.

[13] M. Hu, K. Zeng, Y. Wang, and Y. Guo, "Threshold-based hierarchical clustering for person re-identification," *Entropy*, vol. 23, pp. 522–535, Apr. 2021.

[14] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 7132–7141.

[15] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Munich, Germany, 2018, pp. 3–19.

[16] Z. Zhang, C. Lan, W. Zeng, X. Jin, and Z. Chen, "Relation-aware global attention for person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Seattle, WA, USA, Jun. 2020, pp. 3186–3195.

[17] C.-P. Tay, S. Roy, and K.-H. Yap, "AANet: Attribute attention network for person re-identifications," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 7134–7143.

[18] T. Chen, S. Ding, J. Xie, Y. Yuan, W. Chen, Y. Yang, Z. Ren, and Z. Wang, "ABD-Net: Attentive but diverse person re-identification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Seoul, South Korea, Oct. 2019, pp. 8351–8361.

[19] R. Si, J. Zhao, Y. Tang, and S. Yang, "Relation-based deep attention network with hybrid memory for one-shot person re-identification," *Sensors*, vol. 21, pp. 5113–5133, Jul. 2021.

[20] C. Zhong, G. Qi, N. Mazur, S. Banerjee, D. Malaviya, and G. Hu, "A domain adaptive person re-identification based on dual attention mechanism and camstyle transfer," *Algorithms*, vol. 14, pp. 361–373, Dec. 2021.

[21] X. Wang, H. Zhang, W. Huang, and M. R. Scott, "Cross-batch memory for embedding learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Seattle, WA, USA, Jun. 2020, pp. 6388–6397.

[22] 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)*, Long Beach, CA, USA, Jun. 2019, pp. 598–607.

[23] C. Luo, C. Song, and Z. Zhang, "Generalizing person re-identification by camera-aware invariance learning and cross-domain mixup," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Glasgow, U.K., Aug. 2020, pp. 224–241.

[24] E. Ristani, F. Solera, R. Zou, R. Cucchiara, and C. Tomasi, "Performance measures and a data set for multi-target, multi-camera tracking," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Amsterdam, The Netherlands, Oct. 2016, pp. 17–35.

[25] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable person re-identification: A benchmark," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Santiago, Chile, Dec. 2015, pp. 1116–1124.

[26] D. Gray, S. Brennan, and H. Tao, "Evaluating appearance models for recognition, reacquisition, and tracking," *Perform. Eval. Tracking Surveill.*, vol. 3, pp. 1–7, Oct. 2007.

[27] K. Zeng, M. Ning, Y. Wang, and Y. Guo, "Hierarchical clustering with hard-batch triplet loss for person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Seattle, WA, USA, Jun. 2020, pp. 13657–13665.

[28] 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)*, Seattle, WA, USA, Jun. 2020, pp. 3390–3399.

[29] D. Mekhazni, A. Bhuiyan, G. Ekladious, and E. Granger, "Unsupervised domain adaptation in the dissimilarity space for person re-identification," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Glasgow, U.K., Aug. 2020, pp. 159–174.

[30] 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.*, Salt Lake City, UT, USA, Jun. 2018, pp. 2275–2284.

Y. Hui *et al.*: Unsupervised Cross-Domain Person Re-Identification Method Based on Attention Block and Refined Clustering

IEEE *Access*

[31] Z. Ji, X. Zou, X. Lin, X. Liu, T. Huang, and S. Wu, "An attention-driven two-stage clustering method for unsupervised person re-identification," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Glasgow, U.K., Aug. 2020, pp. 20–36.

[32] Z. Zhong, L. Zheng, S. Li, and Y. Yang, "Generalizing a person retrieval model hetero-and homogeneously," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Munich, Germany, Sep. 2018, pp. 172–188.

[33] W. Xiang, H. Yong, J. Huang, X. Hua, and L. Zhang, "Second-order camera-aware color transformation for cross-domain person re-identification," in *Proc. Asian. Conf. Comput. Vis. (ACCV)*, Nov. 2020, pp. 36–53.

**XIUHUA HU** received the B.S. degree from the Shandong University of Technology, China, in 2010, the M.S. and Ph.D. degrees from Northwestern Polytechnical University, China, in 2013 and 2017, respectively. She is currently a Lecturer with the School of Computer Science and Engineering, Xi'an Technological University. Her current research interests include computer vision, pattern recognition, and intelligent systems.

**YAN HUI** received the B.S. degree from the North China Institute of Technology, China, in 2001, and the M.S. degree from the Lanzhou University of Technology, China, in 2007. She is currently an Associate Professor and the Deputy Director of the Software Engineering Department, School of Computer Science and Engineering, Xi'an Technological University. Her current research interests include computer network management, computer vision, and machine learning.

**HUAN LIU** received the B.S. degree from the Shangqiu Institute of Technology, China, in 2019. She is currently pursuing the graduate degree with the School of Computer Science and Engineering, Xi'an Technological University. Her current research interest includes object tracking.

**XI WU** received the B.S. degree from Heilongjiang University, China, in 2020. He is currently pursuing the graduate degree with the School of Computer Science and Engineering, Xi'an Technology University. His current research interests include computer vision and person re-identification.

**SHIJIE YOU** received the B.S. degree from the China University of Mining and Technology, China, in 2020. He is currently pursuing the graduate degree with the School of Computer Science and Engineering, Xi'an Technological University. His current research interests include computer vision and machine learning.

• • •