Educational Administration: Theory and Practice
2024, 30(6), 692-707
ISSN: 2148-2403
https://kuey.net/
Research Article

Research on Unsupervised Pedestrian Re identification Algorithm Based on Domain Adaptive Method of IBN Network and Label Denoising

Liang Jianbo[1*], Mideth Abisado[2]

[1*] College of Computing and Information Technologies, National University, Manila, 1008, Philippines

[2] School of Big Data Engineering, Kaili University, Kaili,556011, China

***Corresponding author:** Liang Jianbo

[*] College of Computing and Information Technologies, National University, Manila, 1008, Philippines

Funding: Face orientation recognition method based on probabilistic neural network Qiandongnan Ke He J character [2021]46 Research on forest fire real-time detection of UAV platform based on improved YOLOv5 algorithm dongnan Ke He J character [2022]72

Citation:, Liang Jianbo,(2024), Research on Unsupervised Pedestrian Re identification Algorithm Based on Domain Adaptive Method of IBN Network and Label Denoising , *Educational Administration: Theory and Practice*, *30*(6), 692-707,
Doi: 10.53555/kuey.v30i6.5304

ARTICLE INFO

ABSTRACT

This study focuses on the problem of Person Re-Identification (ReID) and deeply explores the application of unsupervised domain adaptation methods in this field. Person re-identification technology is crucial for retrieving specific pedestrian images across cameras and is a core technology for intelligent video surveillance and smart security. Although deep learning has driven significant progress in person re-identification, current methods still face challenges such as scarcity of data annotation and domain differences. To address these issues, we propose a combination of Instance Normalization and Batch Normalization, known as the IBN network module, and a method that utilizes clustering to generate pseudo-labels, thereby implementing person re-identification within an unsupervised domain adaptation framework. We review related technologies and conduct an in-depth study of datasets and evaluation metrics for unsupervised person re-identification. In the experimental section, we validate our approach using the Market-1501 and DukeMTMC-ReID datasets. The results show that both the IBN module and the unsupervised domain adaptation method can effectively improve performance. Additionally, we propose an unsupervised person re-identification method based on label denoising to further enhance accuracy. Although this study has achieved certain results, real-time person re-identification, efficient and robust network design, and the application of unsupervised learning remain focal points for future research.

Keywords: deep learning; unsupervised learning; person re-identification;

1 Introduction

Person Re-Identification (ReID) technology has been a research hotspot in the field of computer vision in recent years. This technology aims to identify and track the movement trajectory of the same pedestrian from different camera perspectives (Li Mingkun, 2023) [1]. This technology has broad application prospects in multiple areas such as intelligent video surveillance, smart security, and smart city construction. However, person re-identification technology still faces many challenges in practical applications (Zhang Hairui, 2023) [2].

The training of deep learning models typically requires a large amount of annotated data. However, in the task of person re-identification, acquiring a sufficient number of accurately annotated datasets is a time-consuming and labor-intensive task (Wang Mingyang, 2023) [3]. In addition, there are significant domain differences in pedestrian images captured from different camera perspectives, meaning that the appearance of the same pedestrian may vary greatly in different environments, posing great difficulties for person re-identification.

To address these issues, unsupervised domain adaptation methods have been introduced into the field of person re-identification. These methods aim to improve the performance of models in new environments using unlabeled data, thereby reducing the reliance on data annotation (Zhang Yang et al., 2024) [4]. However, how to effectively utilize unsupervised learning methods to improve the performance of person re-identification remains the focus of current research.

In this context, this study focuses on the research of unsupervised person re-identification methods. We deeply explore the application of unsupervised domain adaptation methods in person re-identification and propose innovative solutions. By combining the IBN network module, which integrates Instance Normalization and Batch Normalization, with a method that generates pseudo-labels through clustering, we aim to improve the model's adaptability to new environments and enhance the performance of cross-domain person re-identification (Lu Zhiyu, 2023) [5]. Furthermore, we explore an unsupervised person re-identification method based on label denoising to reduce noise in pseudo-labels and further improve the accuracy of person re-identification.

2 Technologies Related to Person Re-Identification

(1) Overview of Person Re-Identification

Person re-identification, also known as Person Re-ID, is a key technology that utilizes computer vision techniques to identify specific pedestrians in images or video streams (Jing Yeyiran et al., 2024) [6]. This technology is often seen as a sub-problem of image retrieval, with the primary task of tracking and retrieving images of the same pedestrian across different devices, based on provided surveillance images of the pedestrian. Person re-identification technology can effectively overcome the field of view limitations of fixed cameras. Additionally, it complements pedestrian detection and tracking technologies, thereby enhancing the performance of the entire surveillance system. Thanks to these advantages, the technology has found widespread applications in key areas such as intelligent video surveillance and intelligent security protection, providing a solid technical backbone for the safety and security of modern society.

(2) Application of Deep Learning in Person Re-Identification

1）Overview of Deep Learning

Deep learning is an emerging research direction in machine learning, driving the rapid development of artificial intelligence. With the popularization of computer science and related disciplines, deep learning has made significant progress in practical applications such as images, videos, and speech (Wang Yunhang, 2024) [7]. Representation learning is a crucial aspect of machine learning, where computers autonomously map data representation to data labels, often requiring manual data annotation and feature extraction. However, deep learning decomposes tasks into multiple subtasks using a divide-and-conquer approach, gradually learning from low-level to high-level representations, automatically completing data representation tasks. Deep learning simulates the operation of the human brain's neural network, consisting of an input layer, intermediate layers, and an output layer, with the intermediate layers performing complex computations (Hu Caixu, 2024) [8].

2）Application of Deep Learning in Person Re-Identification - Convolutional Neural Network (CNN)

In the study of person re-identification, which heavily relies on computer vision techniques, Convolutional Neural Networks (CNN) play a pivotal role. CNN is a widely used architecture in computer vision, capable of performing various tasks such as image classification, segmentation, detection, and recognition (Zhu Songhao et al., 2023) [9]. The structure of a CNN mainly consists of convolutional layers, pooling layers, and fully connected layers, with activation functions performing nonlinear computational processing between layers.

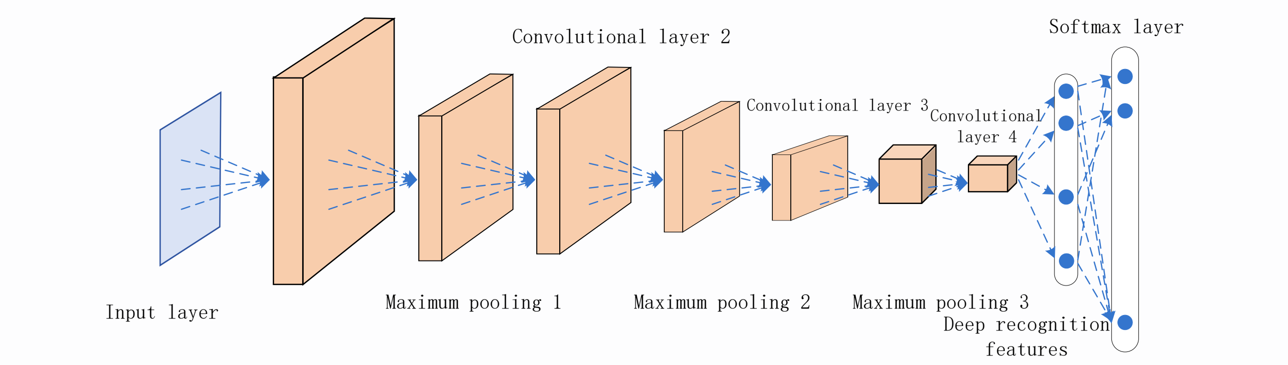

Figure 2.1 Basic structure of DeepID model

Convolutional layers use convolution kernels of various sizes to extract valuable feature information through convolution operations with the image matrix in a template matching manner. Each convolutional layer is typically followed by a pooling layer, which employs average pooling or max pooling techniques to downsample the input feature maps, thereby refining high-dimensional features of the image and simplifying model complexity. The activation layer enhances the model's nonlinear fitting ability through nonlinear functions (Geng Shaosong et al., 2023) [10]. By stacking multiple convolutional layers, pooling layers, and activation layers, the model can extract more advanced image features. Finally, a fully connected layer is added to perform classification or recognition tasks. Figure 2.1 illustrates the structure of the convolutional neural network DeepID, which includes convolutional layers, max pooling layers, and fully connected layers. During training, the network optimizes weight parameters through gradient descent to achieve deep learning for face recognition. To date, there are several classic structures of convolutional neural networks, including LeNet5, AlexNet, VGG, GoogLeNet, ResNet, and DenseNet. To further optimize the network, various algorithms have been introduced, such as Stochastic Gradient Descent (SGD), Momentum SGD, Adaptive Gradient Algorithm (AdaGrad), Root Mean Square Prop (RMSProp), and a combination of the two algorithms called Adam (Yang Donghe, 2023) [11]

(3) Overview of Unsupervised Learning

Unsupervised Learning is one of the important branches of machine learning, along with Supervised Learning and Reinforcement Learning, which together constitute the main methods of machine learning (Zhang Hongwei, 2023) [12]. In Unsupervised Learning, models do not rely on labeled datasets for training but instead discover structures, associations, or patterns by analyzing unlabeled data.

Oord et al. introduced an autoregressive model specifically designed for unsupervised feature learning. They mapped high-dimensional data into a latent variable space and then used autoregressive algorithms to predict future data samples within that space (Zhang Rumeng, 2023) [13]. On the other hand, Hjelm and colleagues focused on improving the effectiveness of representation learning by maximizing mutual information and matching priors. Their proposed DIM (Deep InfoMax) method significantly enhances the applicability of representations in downstream tasks (Zhang Han et al., 2023) [14]. In summary, these studies demonstrate the excellent performance of unsupervised learning in feature extraction and downstream task execution.

3 Unsupervised Person Re-Identification Dataset and Evaluation Metrics

This article focuses on the in-depth study of the application of unsupervised methods in the field of person re-identification. In traditional unsupervised person re-identification methods, domain adaptation techniques are mainly relied on to adjust model parameters, feature dictionaries, and shared subspaces between different domains. A typical approach is to first pre-train using labeled source domain data and then fine-tune the model on unlabeled target domain data through transfer learning (Liu Jingyi et al., 2023) [15]. On the other hand, some scholars have proposed a completely unsupervised method, which directly uses unlabeled target domain data for model training without the need for pre-training on other data. This approach has injected new ideas into person re-identification research

(1) Person Re-Identification Dataset

In this study, two widely used public datasets were selected for in-depth exploration.

Firstly, the Market-1501 dataset, collected at Tsinghua University, consists mainly of pedestrian images of Chinese individuals. It captures 1501 different pedestrian IDs with a total of 32,668 images through six cameras (five high-definition and one low-resolution). Each pedestrian ID has an average of 21 images and comes from at least two different cameras, ensuring data diversity. The dataset is further subdivided into a training set, a test set, and the test set includes a query set and a candidate set. It's worth mentioning that the images in the candidate set are automatically cropped using the Deformable Part Model (DPM) detector. Although the accuracy is slightly lower than manual annotation, it is closer to real-world application scenarios.

Another important dataset is DukeMTMC-ReID, a subset of DukeMTMC collected at Duke University, which mainly contains pedestrian images of Europeans and Americans. The dataset covers 1404 pedestrian IDs with a total of 36411 images, sampled every 120 frames through eight high-definition cameras. Similar to Market-1501, this dataset is also classified in detail. It's worth noting that 408 ID images come from only one camera, adding challenges to person re-identification research. These two datasets, with their unique characteristics, provide valuable resources for in-depth research on person re-identification.

Table 3.1 Dataset parameters

Dataset	Markte-1501	DukeMTMC-ReID
Number of Cameras	6	8
Number of Training Set IDs	751	702
Number of Training Set Images	12936	16522
Number of Test Set IDs	750	702
Number of Test Set Images	19732	17661
Average Number of Images per ID	21	26
Cropping Method	DPM	Manual
Image Size	128X64	Multiple sizes

(2) Evaluation Metrics for Person Re-Identification

The training set in the person re-identification dataset aims to enable the model to learn classification abilities, while the test set is used to verify the model's re-identification performance (Ma, 2023) [16]. Specifically, in the test set, query images are provided through a query set, and a collection of images with the same label is retrieved from a candidate set. In this paper, Cumulative Matching Characteristic curve (CMC) and mean Average Precision (mAP) are adopted as evaluation metrics for person re-identification to assess the model's performance.

CMC Curve: The person re-identification model is evaluated by calculating the rank-n accuracy of the CMC curve. Rank-n accuracy refers to the accuracy of the model when, given a query image from the query set, the pedestrian images in the candidate set are ranked based on feature similarity. Among the top n ranked images, the number of times k the same identity appears is used as a statistic to calculate the accuracy:

$$rank - n = \frac{k}{n}$$

In this paper, rank-1, rank-5, and rank-10 are chosen to represent the CMC curve. The average rank-n values of all N samples in the query set are used as the final evaluation metric:

$$CMC = \frac{1}{N}\sum_{i=1}^{N}(rank - n)_i$$

However, when there are many samples with the same label in the candidate set for a query image, a small value of n cannot accurately evaluate the model's retrieval performance. Therefore, another metric, mAP, is needed for comprehensive evaluation.

Mean Average Precision (mAP): mAP is the overall average of Average Precision (AP) across the query set. In person re-identification, accuracy refers to the proportion of the specified number of target images retrieved to the total number of images retrieved. The AP for a query image is calculated as the average of the Nc samples corresponding to that label in the candidate set:

$$AP = \frac{1}{N_C}\sum_{i=1}^{N_C} p_i$$

Where pi represents the accuracy of retrieving i target images. The mAP can be obtained by averaging the AP of all N query set images, as shown below:

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP$$

A higher mAP value indicates that images of the same identity are ranked higher in similarity sorting, which aligns with the goal of person re-identification.

(3) Unsupervised Domain Adaptation for Person Re-Identification

This paper focuses on the pseudo-label generation method in the target domain for unsupervised domain adaptation in person re-identification. This section will introduce the framework of the model, including the backbone network, clustering method, and loss function.

1）UDA Model Framework

The framework of the unsupervised domain adaptation model for person re-identification is shown in Figure 3.1:

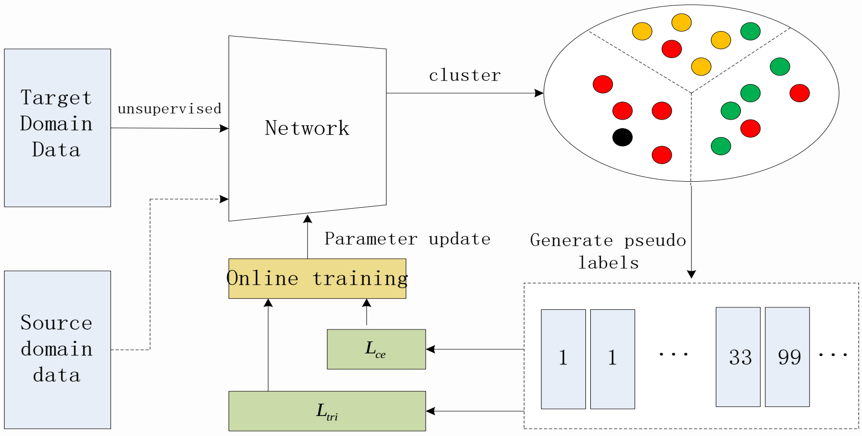

Figure 3.1 Unsupervised Domain Adaptation Model for Person Re-Identification

The model consists of two stages. The first stage involves training on the source domain dataset using the original true labels of the dataset. The network is updated and trained through a loss function, which is supervised learning.

The second stage involves training on the target domain dataset without using the original labels of the dataset. Instead, clustering methods are employed to generate pseudo-labels, which serve as auxiliary information to guide the update and training of the loss function. This stage is unsupervised learning.

The source domain and target domain are not trained independently. Supervised training in the source domain aims to develop a model with a certain pedestrian re-identification capability. This model is then used as a pre-trained model for the second stage, enabling the model to possess a certain pedestrian feature extraction ability at the beginning of target domain training. This guides the clustering algorithm to generate pseudo-labels and perform supervised training.

2）IBN Network Module

Neural networks designed for specific scenarios often suffer from significant performance degradation when applied to other scenarios, mainly due to differences between various contexts. For instance, in the task of person re-identification, if datasets are collected from distinct scenarios, there will be variations in style, color, and other aspects among pedestrian images (Huang Yukun, 2023) [17]. To address this issue, an unsupervised domain adaptation method for person re-identification can be employed. This approach, based on transfer learning, fine-tunes a pre-trained model from one domain (source domain) in another domain (target domain), enabling the network to learn the characteristics of images in the target domain (Wang Wen, 2023) [18]. However, the small batch processing during neural network training can lead to discrepancies in the distribution properties among different batches of data. These differences propagate and amplify through the network layers, potentially affecting the network's convergence speed. To tackle this problem, Batch Normalization (BN) can be used to normalize the input batches, ensuring consistent feature distribution across different batches. On the other hand, Instance Normalization (IN) normalizes individual input data, eliminating variations among instances. By combining BN and IN, the IBN network module is constructed, which preserves both visually invariant features and semantic content features in the shallow layers of the network, while only BN is used in the deeper layers. Specifically, for the task of person re-identification, existing methods often adopt ResNet-50 as the backbone network. In this paper, we propose replacing specific residual modules of ResNet-50 with IBN network modules, forming IBN-ResNet-50, to adapt to person re-identification tasks across different scenarios.

3）Clustering-based Pseudo-label Generation Algorithm

This paragraph explores the application of unsupervised domain adaptation in person re-identification tasks, focusing specifically on strategies that avoid using real labels for training in the target domain. To achieve this goal, this paper proposes a scheme to generate pseudo-labels in the target domain using clustering methods. The paper first outlines two main clustering algorithms: K-Means and DBSCAN. The K-Means algorithm forms K clusters by calculating the distance between samples and cluster centers. However, it has limitations such as the need to manually set the K value, sensitivity to outliers and initial values, and difficulty converging on non-convex datasets. In contrast, the DBSCAN algorithm performs clustering based on density, automatically identifies the number of clusters, adapts to datasets of various shapes, and effectively handles outliers. Nevertheless, it performs poorly on datasets with uneven sample distributions and high-dimensional datasets. A clustering comparison on the Iris dataset demonstrates the characteristics of the two algorithms. In unsupervised person re-identification tasks, where real labels cannot be used and the dataset categories are unknown, this paper opts to use the DBSCAN algorithm for adaptive clustering of the target domain dataset and generating pseudo-labels.

4) Loss Function

Person re-identification is a classification task, and the model's ability to extract pedestrian features directly affects the final re-identification performance (Bai Yangyang, 2023) [19]. In deep learning, neural networks update and optimize network parameters through the backpropagation of gradients from the loss function. This paper will use the cross-entropy loss function and triplet loss function to optimize the parameters of the backbone network for person re-identification.

The cross-entropy loss function calculates the difference between the predicted distribution and the true distribution, helping the model better approximate the true labels. In the multi-class classification task of person re-identification, a multi-class cross-entropy loss function is used: $L_{ce} = \frac{1}{N_b}\sum_{i=1}^{N_b}\sum_{c=1}^{M} -y_{ic}\log(p_{ic})$ And , aiming at the noise and overfitting problems in the dataset, a method of label smoothing for prediction probability is proposed. On the other hand, in order to deal with the misjudgment caused by the high similarity between non similar samples, the triplet loss function is introduced: $L_{tri} = \frac{1}{N_b}\sum_{i=1}^{N_b}\max\left(0, \| f(x_i) - f(x_p)\|_2^2 - \| f(x_i) - f(x_n)\|_2^2 + \alpha\right)$,By optimizing the feature distance between Anchor, Positive, and Negative samples, the classification performance of the model is improved. In addition, an improved triplet loss function is proposed, where θ_{ap} is substituted into the binary cross-entropy loss function. The improved triplet loss function is as follows: $L_{tri} = -\frac{1}{N_b}\sum_{i=1}^{N_b}\log\left[\frac{\exp\left(\| f(x_i) - f(x_n)\|_2^2\right)}{\exp\left(\| f(x_i) - f(x_n)\|_2^2\right) + \exp\left(\| f(x_i) - f(x_n)\|_2^2\right)}\right]$, After training with this loss function, it makes θ_{ap} tend to 0, θ_{an} tend to 1, and the triplet distance gradually tends to: $\| f(x_i) - f(x_n)\|_2^2 << \| f(x_i) - f(x_n)\|_2^2$, Compared to the traditional triplet loss function, this loss function eliminates the need for hyperparameter adjustment of the distance interval a. Joint training is performed using a label-smoothed cross-entropy loss function and an optimized triplet loss function, serving as the overall loss function L for the unsupervised domain adaptation pedestrian re-identification model, $L = L_{ce} + L_{tri}$.

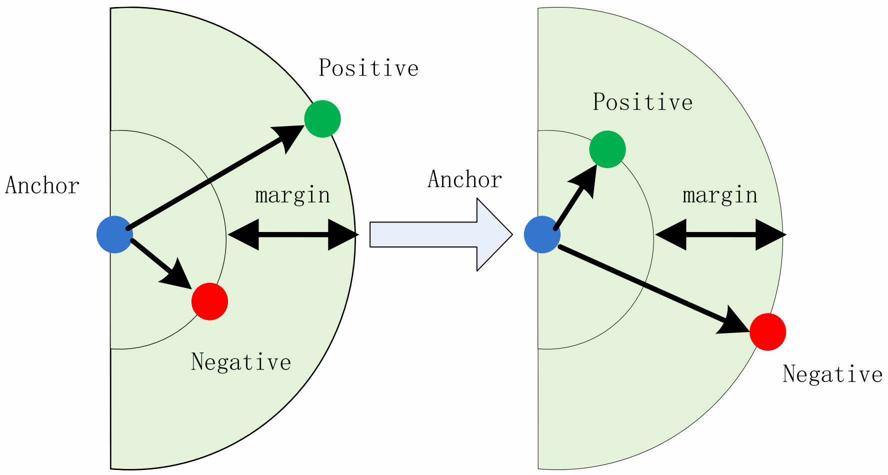

Figure 3.2 Schematic diagram of the triplet loss function

4 Experimental Design and Implementation

(1) Pedestrian Re-Identification Based on Deep Learning

1) Experimental Environment Configuration and Network Training Parameters

The experimental environment configuration includes the use of server hardware such as an Inter Xeon E5-2660 v4 CPU, two NVIDIA GeForce RTX 2080 GPU, server software such as Ubuntu 16.04 system, and the Pytorch 1.1.0 deep learning framework. In the experiment, ResNet-50 and IBN-ResNet-50 were used as backbone networks for comparison. In the supervised training phase of the source domain, a cross-entropy loss function and triplet loss function were adopted, followed by 80 epochs of iterative training. In the unsupervised phase of the target domain, the DBSCAN clustering algorithm was used to generate pseudo-labels to guide the loss function training.

2) Experimental Results Analysis

The experimental results showed that on the Market-1501 and DukeMTMC-ReID datasets, the IBN-ResNet-50 network had significant improvements in various performance indicators compared to the ResNet-50 network. Additionally, the experiments demonstrated that the unsupervised domain adaptation method achieved substantial performance improvements compared to directly transferring the model. Furthermore, the IBN-ResNet-50 network exhibited higher domain adaptation capabilities in cross-domain adaptation experiments. This was primarily attributed to the invariance of the IN layer in the IBN module to changes in image appearance and the ability of the BN layer to preserve content-related information.

The experimental results of unsupervised domain adaptation for pedestrian re-identification using the Market-1501 and DukeMTMC-ReID datasets are presented. The experiment consisted of three stages: supervised training in the source domain, direct transfer of the pre-trained model, and unsupervised domain adaptation training.

In the supervised training phase of the source domain, comparative experiments verified the superiority of the IBN-ResNet-50 network over the ResNet-50 network in pedestrian re-identification tasks, with improvements in various performance indicators. As shown in Tables 4.1 and 4.2:

Table 4.1 Supervised training of Market-1501

methods	mAP	rank-1	rank-5	rank-10
ResNet-50 Source	80.3	92.4	97.2	98.5
IBN-ResNet-50 Source	83.0	93.5	97.5	98.5

Table 4.2 Supervised training of DukeMTMC-reID

methods	mAP	rank-1	rank-5	rank-10
ResNet-50 Source	71.0	83.5	92.2	94.2
IBN-ResNet-50 Source	73.9	86.9	93.7	95.0

During the direct transfer phase of the pre-trained model, it was found that the performance indicators declined significantly when the pre-trained model from the source domain was directly transferred to the target domain for re-identification. As shown in Tables 4.3 and 4.4:

Table 4.3 Direct testing from Market-1501 to DukeMTMC-reID

methods	mAP	rank-1	rank-5	rank-10
ResNet-50 Direct	25.9	41.7	57.0	62.7
IBN-ResNet-50 Direct	32.3	50.1	65.4	70.7

Table 4.4 Direct testing from DukeMTMC-reID to Market-1501

methods	mAP	rank-1	rank-5	rank-10
ResNet-50 Direct	28.3	56.3	72.4	77.9
IBN-ResNet-50 Direct	32.9	62.0	76.6	82.4

In the unsupervised domain adaptation training phase, by using pseudo-labels generated by clustering to guide the training of the loss function, the model performance has been significantly improved compared to direct transfer. As shown in Tables 4.5 and 4.6:

Table 4.5 Domain adaptation training from Market-1501 to DukeMTMC-reID

methods	mAP	rank-1	rank-5	rank-10
ResNet-50 UDA	54.8	70.7	82.9	87.0
IBN-ResNet-50 UDA	61.7	77.0	87.0	90.6

Table 4.6 Ablation experiments from DukeMTMC-reID to Market-1501

methods	mAP	rank-1	rank-5	rank-10
ResNet-50 UDA	65.4	83.3	92.5	94.7
IBN-ResNet-50 UDA	71.5	87.1	94.7	96.6

Through a comparative analysis of the above three experimental results, it was found that the performance indicators of the DukeMTMC-ReID dataset were not as good as those of the Market-1501 dataset. This is mainly because the collection scenario of DukeMTMC-ReID is more complex, including a large number of images obscured by large objects or personnel overlap. These noisy data increase the difficulty of classification, making it difficult for the model to accurately learn pedestrian features. In addition, comparing Experiment 2 and Experiment 3, it was found that the use of unsupervised domain adaptation methods can significantly improve model performance. This is attributed to the fact that this method can help the model adapt to style differences between different datasets and improve generalization ability on unknown target domain datasets, whereas models learned only in the source domain have poor generalization ability. Furthermore, by comparing Experiment 1 and Experiment 3, it was found that using the IBN-ResNet-50 network, which combines the IN layer and BN layer, showed higher performance improvement in cross-domain adaptation experiments. This indicates that the IBN module can enhance the model's domain adaptation ability. Specifically, the IN layer provides invariance to changes in image appearance, allowing the model to be unaffected by cross-dataset style differences, while the BN layer helps preserve content-related information, making pre-trained model parameters more effective.

(2) Unsupervised Person Re-Identification Method Based on Label Denoising

To address the issue of noisy pseudo-labels and improve the performance of person re-identification, an unsupervised person re-identification method based on Label Denoising (LD) is proposed. This method clusters samples based on feature similarity in the pedestrian dataset of the target domain. However, due to limitations in image feature extraction, pseudo-labels often contain noise, affecting subsequent training and re-identification performance. To improve the accuracy of pseudo-labels, various strategies have been adopted in existing studies: such as utilizing representation learning of both global and local information, comparing samples from the target domain and reference domain to generate soft multi-labels, dynamically generating supervisory information for feature representation learning, online correction of noisy pseudo-labels, and improving model performance by bringing similar samples closer and pushing dissimilar samples apart. These strategies aim to reduce the impact of noise and improve the accuracy of person re-identification.

1） Unsupervised Person Re-Identification Model Based on Label Denoising

LD Algorithm Framework: The model adopts a neighbor label correction module to effectively revise the initial pseudo-labels generated by clustering, improving accuracy and reducing noise. Additionally, a class center deviation loss function is introduced to make clusters of samples with the same label more compact. The training process involves pre-training on source domain data, feature extraction from the target domain, clustering to generate initial pseudo-labels, neighbor label correction, and comprehensive optimization through a combination of various loss functions to achieve optimal performance.

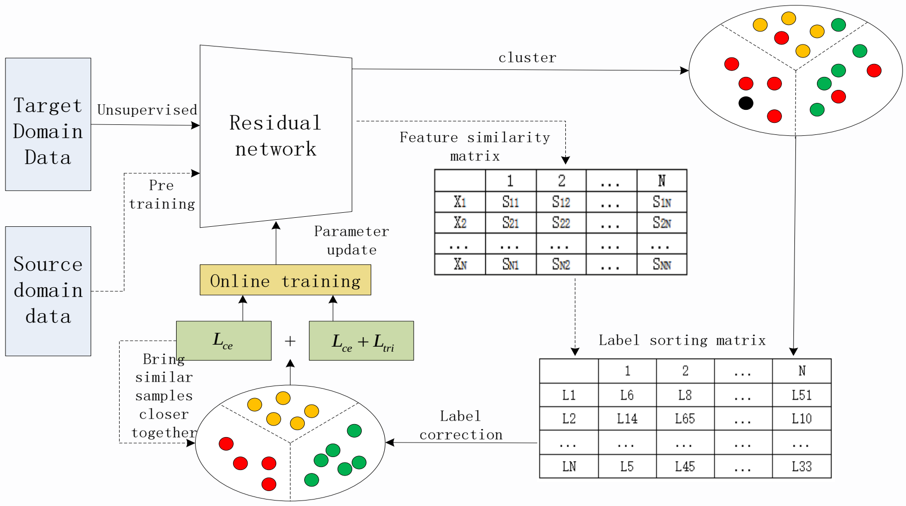

Figure 4.1 Unsupervised Person Re-Identification Model Based on Label Denoising

Nearest Neighbor Label Correction Module: This module, based on the K-nearest neighbor concept, addresses the issue of noisy pseudo-labels caused by domain discrepancies between source and target domain data. It obtains the feature matrix $F_t \in R^{N\times d}$ of target domain data through a feature extraction network and combines it with k-reciprocal coding to generate a re-ranked matrix Dt. Initial hard pseudo-labels are then created using DBSCAN clustering. During the label correction process, Euclidean or cosine distances are used to calculate sample feature distances, which are then used to correct the pseudo-labels, thereby improving the accuracy of person re-identification tasks. This method effectively handles extreme cases that may arise from clustering algorithms, ensuring that pedestrian images of the same identity are more similar at the feature level, which promotes their aggregation into the same cluster, providing reliable support for person re-identification tasks.

Dynamic Neighbor Number and Neighbor Scoring System: The choice of K value in the nearest neighbor label correction module is crucial for pseudo-label correction. Through an analysis of the neighbor situations of target samples after similarity sorting, this study identified three different scenarios: (1) all neighbor samples have the same pseudo-label as the target sample; (2) samples with the same pseudo-label are closer but fewer in number; (3) neighbors include multiple pseudo-labels. In response to these situations, it is recognized that the neighbor situations of different samples vary greatly, so it is necessary to select an appropriate K value based on the situation. Thus, we introduced a dynamically changing neighbor K value and a neighbor scoring system. The dynamically changing neighbor K value is dynamically allocated according to the characteristics of the sample, while the neighbor scoring system aligns the contribution value of neighbor samples with their order or similarity relative to the target sample. Finally, after label correction, the label list needs to be compressed to prevent jump values in the new pseudo-label $y' \in \left[-1, N'_{ct}-1\right]$, ensuring the accuracy and stability of label updates.

Category Center Offset Loss: The key to classification tasks is to cluster samples of the same category in feature space, minimizing the distance between them while maximizing the distance from samples of different categories. To achieve this goal, a center loss function has been proposed and widely used for model optimization. The core of this loss function lies in balancing the minimization of intra-class variation and the separability of features among different classes. Its expression is as follows: $L_c = \frac{1}{2}\sum_{i=1}^{N_b} \| x_i - c_i \|_2^2$, $C_i = \frac{1}{K}\sum_{j=1}^{K} x_{ij}$,This goal is achieved by adjusting the distance between sample features and category center features. However, the traditional center loss function has a limitation: it only focuses on the distance between samples within a batch and cannot guarantee the compactness of the overall cluster. To address this issue, an improved center offset loss function is proposed: $L_{cs} = \frac{1}{N_b}\sum_{i=1}^{N_b} \| x_i - c_i \|_2^2 + \frac{1}{P}\sum_{i=1}^{p} \| c_i - C_i \|_2^2$, $C_i = \frac{1}{N_C}\sum_{j=1}^{N_C} x_{ij}$,It comprehensively considers two aspects: reducing the distance between samples of the same category and reducing the distance between the sample center and the clustering center. Ultimately, the LD model is trained using a combination of cross-entropy loss L_{ce} , triplet loss L_{tri} and center offset loss L_{cs} . The overall loss function L is: $L = L_{ce} + L_{tri} + \lambda L_{cs}$

2）Analysis of Experimental Results

Ablation experiments were conducted on different modules of the proposed label denoising pedestrian re-identification method on two datasets, Market-1501 and DukeMTMC-reID. The experimental results from Market-1501 to DukeMTMC-reID are shown in Table 4.6, and the results from DukeMTMC-reID to Market-1501 are presented in Table 4.7. "Baseline+NLC" represents experiments using the nearest neighbor label correction module with a constant K value. "Baseline+NLC w/ DK" denotes experiments using the nearest neighbor label correction module with dynamic K values and a scoring system. "Baseline+CS" signifies experiments that only added the center offset loss function. LD represents the combination of all modules, which is the proposed label denoising pedestrian re-identification method.

Table 4.6 Ablation experiments from Market-1501 to DukeMTMC-reID

methods	mAP	rank-1	rank-5	rank-10
baseline	61.7	77.0	87.0	90.6
baseline+NLC	63.1	78.1	88.3	91.7
baseline+NLC w/ DK	64.0	78.0	88.1	91.3
baseline+CS	62.2	77.6	86.5	89.8
LD	64.9	79.1	89.2	92.6

Table 4.7 Ablation experiments from DukeMTMC-reID to Market-1501

methods	mAP	rank-1	rank-5	rank-10
baseline	71.5	87.1	94.7	96.6
baseline+NLC	73.4	88.1	95.3	97.2
baseline+NLC w/ DK	74.4	88.9	95.4	96.9
baseline+CS	73.5	88.2	94.8	96.3
LD	74.9	89.0	96.1	97.5

This study conducted an in-depth analysis of the nearest neighbor label correction module and the center offset loss function. The nearest neighbor label correction module refines the labels of target samples based on the principle of feature similarity between pedestrian images of the same category. By adopting dynamic K values and a scoring system, this module significantly improves the performance of re-identification on the Market-1501 and DukeMTMC-ReID datasets. Compared to traditional methods, both mAP and rank-1 metrics have improved. Additionally, to further optimize the distance between intra-class and inter-class samples, a center offset loss function was introduced. Experimental results show that this function effectively reduces the distance between intra-class samples and increases the distance between inter-class samples, positively impacting classification performance.To comprehensively evaluate the overall performance of the proposed method, all modules were combined into the LD model and validated on two datasets. Experimental results indicate that the LD model significantly improves performance compared to the baseline, especially in pedestrian image retrieval tasks, where the LD model retrieves fewer incorrect pedestrian images. This suggests that the nearest neighbor label correction module and the center offset loss function complement each other, jointly enhancing the model's recognition accuracy. By combining the two, the LD model demonstrates excellent performance in pedestrian re-identification tasks.

3）Hyperparameter analysis

Experiments were conducted on important parameters in the LD model, including the dynamically changing K value and the weight λ of the center offset loss function.

The impact of different dynamic K values was investigated. Different numerical settings were applied to the upper limit Ktop and the lower limit Kbottom, while keeping other parameters constant. Experimental results are shown in Tables 4.8 and 4.9. It can be seen that the model achieves the best performance when Ktop is set to 18 and Kbottom is set to 10. This interval more accurately reflects the nearest neighbor situation of the target samples, making the supervision information more accurate and reasonable

Table 4.8 Results of different dynamic K values from Market-1501 to DukeMTMC-reID

methods		mAP	rank-1	rank-5	rank-10
K_{top}=16	K_{bottom}=12	63.8	79.1	87.8	91.1
K_{top}=18	K_{bottom}=10	64.9	79.1	89.2	92.6
K_{top}=20	K_{bottom}=8	63.8	78.4	88.0	91.3

Table 4.9 Results of Different Dynamic K Values from DukeMTMC-reID to Market-1501

methods		mAP	rank-1	rank-5	rank-10
K_{top}=16	K_{bottom}=12	75.4	88.8	95.5	97.3
K_{top}=18	K_{bottom}=10	74.9	89.0	96.1	97.5
K_{top}=20	K_{bottom}=8	74.4	88.3	95.1	97.0

The influence of the coefficient of the center offset loss function. Different values were set for the weight λ of the center offset loss function while keeping other parameters unchanged. Experimental results are shown in Tables 4.10 and 4.11. It can be seen that the model achieved the best performance when λ is set to 1. This indicates that the center offset loss function plays an important role in the method and complements the label correction module.

Table 4.10 Results of Different λ Coefficients from Market-1501 to DukeMTMC-reID

methods	mAP	rank-1	rank-5	rank-10
λ= 0.3	63.4	77.7	87.4	90.4
λ= 0.5	63.9	78.7	88.7	92.1
λ= 0.7	63.1	77.7	87.1	90.4
λ=1	64.9	79.1	89.2	92.6

Table 4.11 Experiments with Different λ Coefficients from DukeMTMC-reID to Market-1501

methods	mAP	rank-1	rank-5	rank-10
λ= 0.3	73.5	87.4	94.9	96.8
λ= 0.5	74.0	88.1	95.2	96.5
λ= 0.7	74.3	88.3	95.5	96.9
λ=1	74.9	89.0	96.1	97.5

4）Comparison with Existing Methods

In this paper, the proposed LD model is compared with existing person re-identification methods based on pseudo-labels in the target domain. The results are shown in Tables 4.12 and 4.13. In the task from Market-1501 to DukeMTMC-reID, the LD model performs second only to MLOL. The advantage of MLOL lies in its training strategy, which reduces the dependence on pseudo-labels by deeply exploring the relationship between target domain data, thereby improving performance. In the task from DukeMTMC-reID to Market-1501, the LD model performs second only to LNL. LNL adopts a strategy of mutual correction between two models, which is superior in performance to the LD model that only uses a single model. Overall, the LD model demonstrates strong competitiveness in both tasks.

Table 4.12 Comparison of Results between Market-1501 to DukeMTMC-reID and Existing Methods

methods	mAP	rank-1	rank-5	rank-10
BUC	27.5	47.4	62.6	68.4
MAR	48.0	67.1	79.8	84.2
SSG	53.4	73.0	80.6	83.2
AD-Cluster	54.1	72.6	82.5	85.5
CVSE	56.1	75.3	82.9	85.4
LNL	62.5	77.4	88.1	90.6
MLOL	69.8	83.1	90.8	93.0
LD (ours)	64.9	79.1	89.2	92.6

Table 4.13 Comparison between DukeMTMC-reID to Market-1501 and the results of existing methods

methods	mAP	rank-1	rank-5	rank-10
BUC	38.3	66.2	79.6	84.5
MAR	40.0	67.7	81.9	87.3
SSG	58.3	80.0	90.0	92.4
AD-Cluster	68.3	86.7	94.4	96.5
CVSE	63.2	84.1	92.8	95.0
SPCL	73.1	88.1	95.1	97.0
LNL	75.2	90.9	96.4	97.9
MLOL	70.9	86.6	93.1	95.1
LD (ours)	74.9	89.0	96.1	97.5

5 Conclusion and Prospect

Person re-identification is a task aimed at retrieving specific pedestrians across non-overlapping camera views. It finds widespread applications in suspect tracking, missing person searches, and more. With the popularization of deep learning, research on person re-identification has further advanced. However, due to factors such as the diversity of pedestrian postures, variations in camera angles, and resolution differences, person re-identification remains a challenging task.

This study delves into unsupervised domain adaptation for person re-identification. By introducing the IBN network module and generating pseudo-labels through clustering, the cross-domain recognition performance of the model between different datasets has been significantly improved. Experimental results show that the IBN-ResNet-50 network structure achieves higher mAP and rank-n accuracy on both the Market-1501 and DukeMTMC-ReID datasets compared to the traditional ResNet-50 network, verifying the effectiveness of the IBN module in enhancing the model's domain adaptation ability. Additionally, through unsupervised domain adaptation training, the model can learn from unlabeled data in the target domain, effectively addressing the scarcity of labeled data in real-world applications.

Furthermore, this study proposes an unsupervised person re-identification method based on label denoising (LD model). This method effectively reduces noise in pseudo-labels through a nearest neighbor label correction module and a class center deviation loss function, further improving the accuracy of person re-identification. Ablation experiments and hyperparameter analysis validate the effectiveness of each module and determine the optimal hyperparameter configuration. Comparative experiments with existing methods demonstrate that the LD model achieves competitive performance in cross-domain person re-identification tasks.

Despite the achievements made in this study in the field of unsupervised domain adaptation for person re-identification, numerous challenges and future research directions remain. Firstly, with the increasing number of surveillance cameras, effectively processing massive amounts of surveillance video data to achieve real-time person re-identification will be a crucial direction for future research. Secondly, as deep learning technology continues to evolve, designing more efficient and robust network structures for person re-identification to improve the model's generalization ability across different scenarios is also an important topic for future study. Additionally, unsupervised learning, as a current research hotspot in the field of machine learning, holds promise for better applications in person re-identification, especially in addressing the challenge of pseudo-label noise.

References

1. Li Mingkun (2023) Unsupervised pedestrian re identification effective feature learning [PhD thesis, Beijing University of Posts and Telecommunications] DOI:10.26969/d.cnki.gbydu.2023.000309
2. Zhang Hairui (2023) Research on pedestrian re identification technology and embedded implementation [Ph.D. thesis, Nanjing University of Posts and Telecommunications] DOI:10.27251/d.cnki.gnjdc.2023.001516
3. Wang Mingyang (2023) Research on pedestrian re identification algorithm based on deep learning [doctoral thesis, Heilongjiang University] DOI:10.27123/d.cnki.ghlju.2023.002019
4. Zhang Yang, Wang Ronggang,&Lv Ming (2024) Intelligent security system based on pedestrian re identification Industrial control computer, 37 (04), 99-100
5. Lu Zhiyu (2023) Research on Unsupervised Pedestrian Re identification Algorithm and Its Cross modal Application [Doctoral Dissertation, Nanjing University of Posts and Telecommunications] DOI:10.27251/d.cnki.gnjdc.2023.001593
6. Jing Ye Yi Ran, Yu Zeng, Shi Yun Xiao, etc (2024) A review of research on unsupervised adaptive pedestrian re identification based on pseudo labels Computer Science, 51 (01), 72-83
7. Wang Yunhang (2024) Research on Traffic Sign Detection and Recognition Methods Based on Deep Learning [J/OL] Automotive Digest, 1-9 https://doi.org/10.19822/j.cnki.1671-6329.20240075
8. Hu Caixu (2024) Research on automatic recognition technology of urban green spaces based on oblique photography and deep learning Shanxi Architecture, 50 (11), 190-194 DOI:10.13719/j.cnki.1009-6825.2024.11.045
9. Zhu Songhao,&L ü Zhihan (2023) A cross modal pedestrian re identification method based on graph convolution Journal of Nanjing University of Posts and Telecommunications (Natural Science Edition), 43 (02), 53-62 DOI:10.14132/j.cnki.1673-5439.2023.02.007
10. Geng Shaosong,&Li Jinguo (2023) Multi feature fusion pedestrian re recognition based on dynamic convolution and attention Computer Engineering and Design, 44 (04), 1228-1234 DOI:10.16208/j.issn1000-7024.2023.04.036
11. Yang Donghe, Ren Guoyin,&Zhang Xiaolin (2023) Research on unsupervised attention mechanism for pedestrian re identification based on group convolution [J] Journal of Inner Mongolia University of Science and Technology, 42 (02), 128-133 DOI:10.16559/j.cnki.2095-2295.2023.02.006
12. Zhang Hongwei (2023) Research on unsupervised pedestrian re identification method using contrastive learning [D] Nanjing University of Information Technology DOI:10.27248/d.cnki.gnjqc.2023.001627
13. Zhang Rumeng (2023) Research on Semi supervised Pedestrian Re Recognition Based on Single Camera Annotation [D] Shandong Normal University DOI:10.27280/d.cnki.gsdsu.2023.001069
14. Zhang Han, Zhao Kai,&Wang Kefeng (2023) Design and implementation of cross lens pedestrian re identification system [J] Information and Computer (Theoretical Edition), 35 (16), 154-159

15. Liu Jingyi, Jin Bin, Jie Xiangxin, etc (2023) Research on pedestrian re recognition based on attention scoring function [J] Wireless Internet Technology, 20 (12), 131-133
16. Ma Ru (2023) Research on pedestrian re identification algorithm combining skeleton information [D] Inner Mongolia University of Science and Technology DOI:10.27724/d.cnki.gnmgk.2023.000902
17. Huang Yukun (2023) Research on pedestrian re identification methods for degraded images [D] University of Science and Technology of China DOI:10.27517/d.cnki.gzkju.2023.000852
18. Wang Wen (2023) Multi dimensional information joint modeling and representation for pedestrian re identification [D] Beijing Jiaotong University DOI:10.26944/d.cnki.gbfju.2023.002419
19. Bai Yangyang (2023) Research on Domain Adaptive Pedestrian Re identification Algorithm for Monitoring Scenarios [D] Nanjing University of Information Technology DOI:10.27248/d.cnki.gnjqc.2023.001041