Unsupervised Multi-source Adaptive Pedestrian Re-recognition: Based on Target Domain Prioritization and Multi-dimensional Edge Features

Jia He[1,2], Xiaofeng Zhang[1(✉)], Tong Xu[3], Mingchao Zhu[1], Kejun Wang[1,3], Pengsheng Li[1], and Xia Liu[1]

[1] Beijing Institute of Technology, Zhuhai, Zhuhai 519000, Guangdong, China
karen6886@163.com
[2] City University of Macau, Macau 999078, China
[3] Harbin Engineering University, Harbin 150000, Heilongjiang, China

Abstract. Unsupervised multi-source domain adaptation facilitates the transfer of knowledge from multiple source domains, which possess labeled data, to an unlabeled target domain. Pedestrian re-identification is a technique for cross-camera pedestrian retrieval in surveillance data. The utilization of multiple source domains holds significant research implications, particularly in scenarios involving a substantial volume of data. Recently, efforts have been made to eliminate distributional differences between data from different domains. However, these approaches do not take into account the specificity of the target domain. Furthermore, when using graph convolutional networks for domain fusion, few studies have explored the utilization of deep correlations between nodes, which is crucial for node updates in GCNs. In this paper, we present a novel methodology that enhances domain fusion and showcases robust performance. In particular, we first propose a Domain Fusion Module based on the prioritization of target domains, which enables the fusion of domain information. Second, in order to mine the deep correlation between nodes, we process it by introducing multidimensional edge features. Our multiple experimental results using four datasets on three migration tasks demonstrate the superior performance of the DFTM, thus providing strong support for the effectiveness of our approach.

Keywords: multi-source domain adaptation · pedestrian re-identification · domain integration

1 Introduction

Pedestrian re-recognition is a technique for cross-camera pedestrian retrieval in surveillance data. Given an image, the task of pedestrian re-recognition involves identifying a pedestrian image that corresponds to the identity of the query image from a collection of multiple candidate images captured by diverse devices. This process, in turn, enables the

V. C. M. Leung et al. (Eds.): Qshine 2023, LNICST 574, pp. 315–329, 2024.
https://doi.org/10.1007/978-3-031-65123-6_23

automatic retrieval of the target pedestrian. Scholars have extensively and thoroughly explored this area, and they have made significant progress in achieving high accuracy and speed in pedestrian re-recognition by applying deep learning techniques [24, 25]. Currently, the use of deep learning with labeled datasets (source domain) has shown promising results [1–3]. However, the performance drops drastically when applying this model to new scenarios (target domain) due to the presence of domain bias between different scenarios. Additionally, the cost incurred in labeling the dataset is high, making unsupervised domain adaptation methods [garnering] more and more attention.

Domain adaptation facilitates the transfer of knowledge acquired from the labeled source domain to the unlabeled target domain. Presently, the state-of-the-art methodology for pedestrian re-identification relies on pseudo-label generation in the target domain [4–6]. The trained model is initially obtained through training on the labeled source domain.Then, the target domain is iteratively fine-tuned by pseudo-label prediction. The common feature of these studies is that only one source domain is used for model pre-training. However, in real applications, we often have data from multiple domains, making multi-source domain adaptation (MSDA) pedestrian re-identification [7] more valuable to study. For unsupervised multi-source domain adaptation, the improvement brought by expanding the dataset by simply superimposing different source domains is limited or even negative. Since the data comes from separate domains, the core challenge of unsupervised multi-source domain adaptation lies in eliminating the distributional differences between the data from these different domains. Previous related studies [7, 8] have typically used graph convolutional networks to address this challenge. This involves treating features from different domains as nodes of a graph, facilitating feature interaction between nodes through the information transfer mechanism of graph convolutional networks, and thus achieving domain fusion. Although such a mechanism allows for some degree of domain fusion, there still remains the problem of incomplete fusion and the underutilization of deep correlations between nodes.

In this paper, we propose a new framework, as shown in Fig. 1, that better reduces domain differences and leverages a method to mine higher-order correlations of neighboring nodes. Our approach incorporates two new techniques. First, we introduce a Domain Fusion based on Target domain prioritization and Multidimensional edge features module (DFTM). While Graph Convolutional Networks can be employed for domain fusion in previous studies [7, 8], our DFTM module prioritizes the dissimilarity between the distributions of the target domain and the source domain, while considering the similarity of distributions within the source domain. Through considering the source domain as a cohesive entity and disregarding the fusion among its individual components, we place greater emphasis on the fusion between the target domain and the source domain. This approach is referred to as target domain prioritization. Secondly, DFTM effectively mines the high-order correlations of neighboring nodes, and we utilize multidimensional edge features to deeply explore the correlation information between nodes, enhancing the diversity of node updates.

We conducted three distinct migration experiments using four well-established datasets. The obtained results showcase the method's noteworthy performance. In summary, our contributions are as follows:

1. We propose DFTM, which fuses features from different domains and narrows the gap between them.
2. We introduce a graph convolutional network node updating method that incorporates multidimensional edge features to enable comprehensive mining of relevant information.

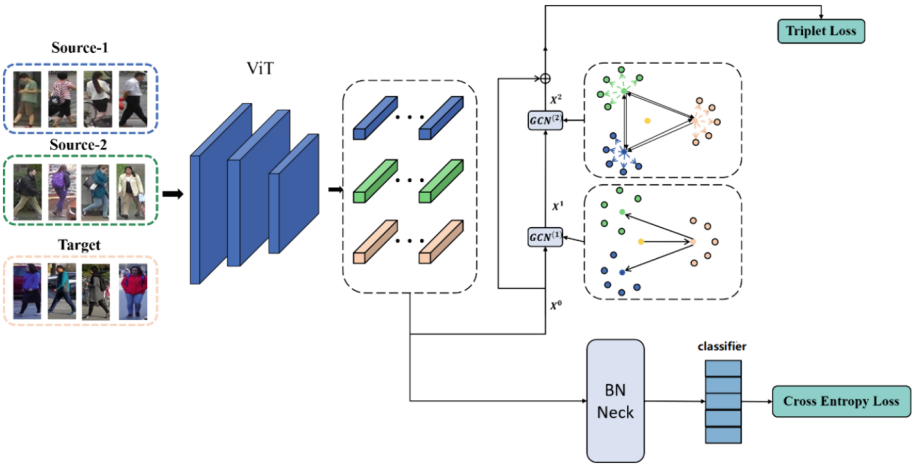

Fig. 1. Overview of our proposed framework, comprising three main components: (1) A feature extractor responsible for extracting features from various domains. (2) The Domain Fusion with Target domain prioritization Module (DFTM), devised to enable message passing and feature fusion across diverse domains. The color consistency denotes the same domain. (3) The extracted features are divided into two branches to calculate the loss function: one branch computes the Cross Entropy Loss after BN Neck, and the other branch is fed into the DFTM for fusion and subsequently utilized to compute the Triplet Loss.

2 Related Work

2.1 Unsupervised Domain Adaptation for Pedestrian Re-recognition

In the context of unsupervised domain adaptation for pedestrian re-recognition, alignment between the source and target domains is imperative owing to substantial distributional disparities between the two domains. In single-source domain adaptation, initial methodologies employed image-based domain migration to align samples from the target domain with the source domain. These approaches include SPGAN [9], PTGAN [10], and AD-Cluster [11], among others. More recent approaches, on the other hand, focus on feature dimensionality and aim to extract domain invariants for all domains. For instance, Mekhazni et al. [12] proposed the D-MMD method, which designs a disparity-based MMD loss for performing inter-domain feature alignment. Huang et al. [13] introduced the DAAM method, utilizing attention to extract domain-shared features and a domain similarity loss to aid the domain adaptation process. Yang et al. [14] presented the PAUL method, which achieves fine-grained domain alignment by learning

patch-level features. Additionally, Tao et al. [15] proposed the MGML method, which not only aligns domains in the patch dimension but also considers distributional differences between cameras, using the G-MCD loss to simultaneously improve inter-domain bias and intra-domain variations.

2.2 Unsupervised Multi-source Domain Adaptation for Pedestrian Re-recognition

For unsupervised multi-source domain-adapted pedestrian re-recognition, there is still a lack of sufficient studies, and the core problem lies in addressing the inter-domain discrepancy. Currently, domain fusion stands out as one of the more effective methods. Among the previous domain fusion approaches, Qi et al. proposed DSAF [16], which involves setting a unique network path for each source domain and adaptively mitigating the domain gap between multiple source domains and the target domain through an additional path to achieve fusion. Wang et al. introduced MSTNet [17], achieving sample dimension fusion based on generative confrontation and feature dimension fusion through distribution learning. The most promising method is MDIF based on graph convolutional networks [7], which initially extracts features from samples of different domains using a basic feature extraction network, and then fuses these features through a two-stage GCN. However, these methods fail to consider the specificity of the target domain during domain fusion, and they also lack the utilization of the deep correlation between nodes. As a remedy, we propose a domain fusion mechanism that prioritizes the target domain, as well as a method that thoroughly mines the correlation between nodes.

2.3 Graph Convolutional Networks

The module designed in this paper is based on graph convolutional networks (GCNs), which are convolutional neural networks proposed by researchers for processing graph data structures, drawing on the structure and ideas of conventional convolutional networks. Graph Convolutional Networks can be broadly categorized into two main types: spectral domain-based approaches and spatial domain-based approaches. Spectral domain-based methods were first proposed by Bruna et al. [24], followed by the introduction of GCN by Thomas et al. [25], which utilizes spectral graph convolution with local first-order approximation to process graph data. Defferrard et al. [26] and others introduced ChebNet, which uses Chebyshev polynomials instead of convolution kernels in the spectral domain to speed up the time of convolution operations. Spatial domain-based approaches are now more mainstream in academia due to their greater flexibility and intuitiveness compared to spectral domain-based approaches. DCNN [27], for instance, samples 1 to k neighbors to obtain a representation of a node. Hamilton et al. [28] proposed GraphSAGE, a general-purpose generalization framework to efficiently generate node embeddings of previously unseen data using node feature information. Velickovic et al. [29] introduced GAT, which utilizes a self-attention mechanism in aggregating neighboring node representations, enabling the network to learn the weights of different neighboring node representations. Xu et al. [30] proposed Graph Isomorphism Network (GIN), which approximates graph isomorphism tests of the mapping function by fusing the neighboring features and processing the original features using a fully connected

layer with strong fitting capabilities.Building upon the limitation that previous methods treat edges as mere binary values (0/1), indicating only whether the nodes are connected or not, but not capturing deeper connections between nodes, Gong et al. [31] proposed EGNN, which enables the model to learn the deeper relationships between nodes and simultaneously update the node features with multidimensional edge features. With the inclusion of multi-dimensional edge features, finer-grained information transfer between nodes can be achieved, leading to improved results.

3 Methodology

3.1 Overall Structure of DFTM

The module consists of two Graph Convolutional Networks (GCNs). The first GCN is responsible for bridging the gap between the source and target domains. The second GCN establishes unidirectional connections from each node to the center node of the domain and bidirectional connections between the center nodes of different domains. These connections enable the domain center nodes to exchange information with each other, facilitating domain fusion.

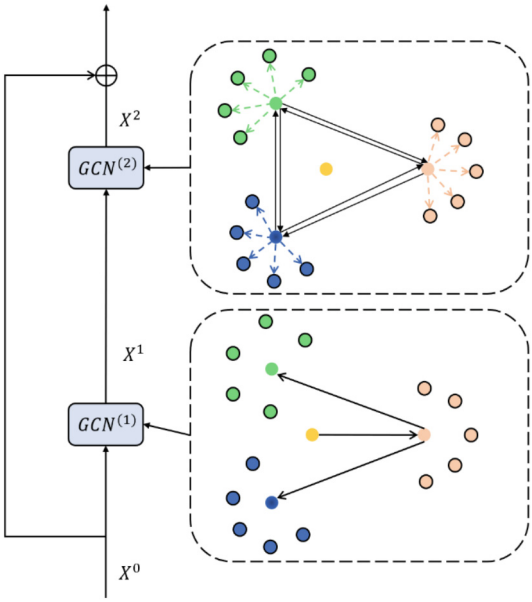

Fig. 2. This figure shows our proposed DFTM module consisting of two GCNs. Blue and green represent the source domain sample nodes, brown-pink represents the target domain sample nodes, blue, green, and total pink without black outlines represent the center nodes of each domain, yellow represents the center nodes of all source domains, dashed lines represent the edges within the domains, and solid lines represent the edges between the domains.

3.2 Domain Fusion with Target Domain Prioritization Module

To address the problem of poor discriminability in target domain features, which results in a significant distribution difference compared to other domains, this section proposes a novel domain fusion mechanism that prioritizes target domains. The entire mechanism is illustrated in Fig. 2. For simplicity and without loss of generality, we assume there is one source domain and two target domains, each containing a set of samples. Specifically, the domain fusion mechanism can be divided into two stages.

The first stage is shown in Fig. 2. GCN$^{(1)}$ A domain center node is maintained for each domainagt that is used as a representative of that domain. The calculation is done by weighting all the nodes of the domain in a batch with the following formula:

$$agt^d = \sum_{n=1}^{Q} w_n^d x_n^d \tag{1}$$

where agt^d represents the central node of the d-th domain, Q denotes the number of samples belonging to the d-th domain in a mini-batch. The weights $w_n^d = f\left(x_n^d\right)$ are learned from the samples x_n^d and are normalized by $w_n^d = w_n^d / \sum_{m=1}^{Q} w_m^d$ to ensure that the sum of weights is equal to 1, where $f(\cdot)$ represents a fully connected layer. Through such weighted averaging, the central node of domain d can be obtained. However, this method operates only within a mini-batch, making it difficult to achieve overall coverage of domain d and lacking globality and robustness. Therefore, in this section, a sliding exponential average is used to maintain a robust domain central node, as shown in Eq. (2):

$$(\ddot{agt}^d)^t = (1-\alpha)\,(\ddot{agt}^d)^{t-1} + \alpha\left(agt^d\right)^t \tag{2}$$

where t represents the t-th iteration, and α denotes the magnitude of the update parameter. Through the aforementioned sliding exponential average updates, a domain central node covering the entire domain d can be obtained, thereby avoiding the network from getting trapped in local optima.

After obtaining the domain center node, the method further distinguishes between the source and target domains. As shown in Fig. 2, the domain center nodes of all source domains are first weighted and averaged to obtain a source domain center node that can cover all source domains, denoted as agt_{source}, specifically shown in Eq:

$$agt_{source} = \sum_{d=1}^{D} a_d agt^d \tag{3}$$

where D represents the number of source domains, and $a_d = f'\left(agt^d\right)$ is a learned weight value, which is similarly normalized by $a_d = a_d / \sum_{d=1}^{D} a_d$. Normalization is performed to ensure that the sum of the weights is 1. Here, $f'(\cdot)$ represents a fully connected layer. By weighting all the source domain center nodes as described above, it is possible to obtain a center node that represents the entire source domain, denoted as agt_{source}, and these nodes are presented as light yellow nodes in the graph.

After obtaining the center node, the construction of the first layer of the directed graph convolutional network is carried out, and the specific connections are illustrated in Fig. 2 as $GCN^{(1)}$. In this network, the center node of the source domain points to the center node of the target domain, while the center node of the target domain points to the center nodes of the two source domains, respectively. This connection allows the center node of the target domain to receive information from the source domain as a whole and also allows the center nodes of the two different source domains to receive information from the target domain. As a result, the goal of bringing the source and target domains closer to each other is achieved. This approach prioritizes addressing the significant distribution difference between the target domain and the source domain, while considering the relatively smaller distribution differences among the source domains themselves. Consequently, the source domain is initially treated as a collective entity, overlooking the fusion of individual source domains and instead focusing on the fusion of the target domain with the source domain, i.e., prioritizing the target domain. Through the first layer of the graph convolutional network, the distance between the source and target domains can be reduced.

In the second stage of fusion, as in Fig. 2 $GCN^{(2)}$ shown, a unidirectional connection is first constructed for each node with the center node of the domain, and a bidirectional connection is constructed between the domain center nodes. Through such connections, the domain center nodes are able to receive information from each other and achieve domain fusion. Individual nodes can also receive information from other domains indirectly by connecting with the center node. Eventually, the domains become closer to each other.

To summarize, in the first stage, the distance between the source and target domains is reduced by using $GCN^{(1)}$ to bring the source and target domains closer to each other, while in the second stage, by $GCN^{(2)}$ achieving mutual proximity between all domains. Through this two-stage fusion mechanism, it is possible to pay better attention to the target domains, thus ensuring a stronger fusion of the target domains, which is more in line with the requirements of the unsupervised multi-source domain adaptation to the pedestrian re-recognition task.

3.3 Multi-dimensional Edge Feature Based Graph Convolution Node Update Method

In the previous section, a two-stage domain fusion mechanism was proposed based on a two-layer graph convolutional network, and the connectivity of the network was given. In this section, a message passing method is constructed for graph convolutional networks to capture higher-order correlations of neighboring nodes. Specifically, in a regular graph convolutional network, the edges are mere $0/1$ structures, indicating only whether the nodes are connected or not, without providing information on how the nodes are connected to each other or their correlations. However, in graph structures, nodes are not simply adjacent or non-adjacent to each other; there is more information to be explored. Therefore, in this section, edges are represented as multi-dimensional features. For example, if node v_i is adjacent to node v_j and the directed edge is from v_j pointing to v_i, then there exists an edge feature $e_{i,j} \in \mathbb{R}^{P\times 1}$, where an edge feature is a vector of

length P. Its generation is obtained by learning from a multilayer perceptron, as shown in Eq:

$$e_{i,j} = \mathrm{MLP}([v_i, v_j]) \tag{4}$$

where $[v_i, v_j]$ denotes the transformation of a node feature of length D for the node features v_i and v_j, which are spliced into a new vector of length 2D. Through a two-layer Multilayer Perceptron (MLP), the output is a multidimensional vector of length P. This vector contains multidimensional edge features of length $e_{i,j}$, which represent the depth-dependent information between v_j and v_i when pointing.

Based on the obtained multidimensional edge features, we can implement the following message passing and node updating methods, taking node v_i as an example:

$$\hat{v}_i = v_i + g(v_i) \tag{5}$$

$$g(v_i) = \mathrm{W}_e \sum_{v_k \in \mathcal{N}_{v_i}} \left(e_{i,k} \circ v_k\right) \tag{6}$$

where $\mathcal{N}_{v_i}$ is the set of neighboring nodes of node v_i, i.e., v_k is the set of neighbor nodes of node v_i, and there exists v_k pointing to node v_i with edge features that point to $e_{i,k}$. The symbol $\circ$ represents broadcast multiplication, and $W_e \in \mathbb{R}^{1\times P}$ is a learnable weight used to fuse the P features of individual channels. Through the message passing mechanism $g(\cdot)$, we obtain the node after receiving information from neighboring nodes $g(v_i)$ and update the node features by the residual join with v_i. The updated node features are obtained through residual linking. By using the multidimensional edge features $e_{i,k}$, the message passing mechanism can aggregate information from P different channels separately, allowing for diverse node updates and mining of more correlations. In the two-layer graph convolutional network constructed in this paper, all nodes are updated using the aforementioned method.

4 Experiment

4.1 Data Set and Experimental Setup

We utilized four datasets, namely Market1501 [18], DukeMTMC [19], CUHK03 [20], and MSMT17 [10]. The analysis of the CUHK03 dataset in [21] served as a reference and was used solely as the source domain, while the other datasets were taken in turns to act as the target-domain data. Market1501 [18] comprises a total of 32,668 labeled pedestrians with 1501 unique identities. DukeMTMC-ReID [19] includes images of 1852 individuals spread across 8 cameras, and among them, 1413 pedestrians appeared in multiple cameras. CUHK03 [20] contains 13164 images encompassing 1360 different pedestrian identities. MSMT17 [10] consists of a total of 126,441 images with 4101 pedestrian identities.

Vision Transformer was used as the base model for feature extraction, employing data enhancement schemes including resizing, random horizontal flipping, padding, random cropping, random erasing, and clipping and splicing to fully augment the experimental

data. The batch size for training was set to 128, meaning that each domain was assigned 32 samples. Stochastic small-batch gradient descent with first-order momentum was used as the optimizer, and the initial learning rate was set to 0.003 using a multi-step learning rate tuning scheme, where the learning rate decayed to one-tenth of the original at the 40th and 70th epochs, respectively. For the hyperparameter α used in the sliding exponential mean update during this experiment, it was set to 0.1. The MLP used to generate multidimensional edge features with two layers utilizes GeLU as the activation function. Regarding evaluation metrics, Rank-1 and MAP are used to measure the performance of the proposed domain fusion method.

4.2 Performance Comparison Experiment

We conducted experiments using the DFTM method on various migration tasks and provided a comprehensive and intuitive analysis of its performance by comparing it with previous related methods. Firstly, we compared it with a related domain fusion scheme for the unsupervised multi-source domain adaptation pedestrian re-identification task to demonstrate its effectiveness. Additionally, as there are limited methods related to unsupervised multi-source domain adaptation for pedestrian re-recognition, we included single-source domain adaptation methods in the comparison as well. However, direct comparison between multi-source and single-source domain adaptation methods is not reasonable due to the data expansion with multi-sources. Therefore, we selected the single-source domain adaptation method with the best result in the migration task to perform the comparison. For example, when comparing the migration task → Market-1501, the single-source domain adaptation method will be compared to the single-source domain adaptation method in DukeMTMC → Market-1501, CUHK03 → Market-1501, CUHK03, etc. and select the optimal one as its result. While multi-source domain adaptation takes only the migration tasks (DukeMTMC, CUHK03, MSMT17) → Market-1501's experimental results.

We compared a total of three types of domain fusion and domain alignment methods, which are Image Style Transfer (IST) [9–11] - converting the source domain to the target domain in the image dimension, Feature Alignment (FA) [12–15] - aligning the source and target domains in the feature dimension to obtain domain invariants, and Multi-source Domain Adaptation (MDA) [7, 16] - based on unsupervised multi-source domain adaptation. The primary experimental findings are presented in Table 1.

For the method DFTM proposed in this paper, it achieves the best results on the task of migrating to Market-1501, exceeding the previous multi-source domain adaptation method MDIF on Rank-1 metrics 0.6% and improves on the mAP 0.7%. On the task of migrating to DukeMTMC, the DFTM method also achieves optimality, with an improvement in Rank-1 compared to MDIF, and an improvement in mAP 1.8% improvement over MDIF on Rank-1, and an improvement on mAP of 1.3% on Rank-1 and on mAP. On the MSMT17 migration task, DFTM is only slightly lower in Rank-1 compared to the currently optimal method MGML 0.9% on Rank-1 compared to the current optimal method MGML.

In conclusion, our experiments on diverse migration tasks substantiate the effectiveness of the proposed Domain Fusion with Target domain prioritization Module (DFTM)

Table 1. Comparison of method results on different migration tasks

Type	Method	→ Market-1501		→ DukeMTMC		→ MSMT17	
		Rank-1	mAP	Rank-1	mAP	Rank-1	mAP
FA	PAUL [14]	68.50	40.1	72	53	34.3	18.9
	D-MMD [12]	70.6	48.8	63.5	46.0	34.4	15.3
	DAAM [13]	86.4	67.8	77.6	63.9	46.7	21.6
	MGML [15]	88.3	68.5	79.9	64.4	49.1	23.0
IST	PTGAN [10]	38.6	–	27.4	–	11.8	–
	SPGAN [9]	58.1	26.9	46.9	26.4	–	–
	AD-Cluster [11]	86.7	68.3	72.6	54.1	–	–
MDA	DSAF [16]	79.7	53.2	66.2	40.1	42.4	16.8
	MDIF [7]	92.1	79.4	79.1	65.0	47.7	21.3
	DFTM	92.7	80.1	80.9	66.3	48.2	23.8

method, which surpasses other approaches and attains superior performance on multiple migration tasks.

4.3 Ablation Experiments

The DFTM method consists of two main components, so these two components were split and ablation experiments were performed separately. To facilitate a comprehensive comparison of the methods, these two components are juxtaposed with other related approaches to evaluate their respective strengths and weaknesses. Specifically, the target domain prioritization mechanism's essence lies in the establishment of connections solely between the source and target domains through the initial layer of the graph convolutional network. Thus it is compared to a fully connected graph convolutional network (connections are also constructed between the source domains). For multidimensional edge features, it is compared to the underlying 0/1 edge structure is compared. For simplicity, the domain fusion mechanism of Target Domain Priority is replaced by TDP (Target Domain Priority), Non-Target Domain Priority (fully connected form) is replaced by NTDP (Non-Target Domain Priority), multidimensional edge features are replaced by High, and the underlying 0/1 edge structure is replaced by Basic. The results of the ablation experiments are shown in Table 2.

As shown in the table, there is an improvement in the effectiveness of the model after the introduction of a simple non-target domain prioritized fully connected fusion mechanism. When the target domain prioritized fusion mechanism is used, there is a further significant improvement in the performance of the network, which proves the effectiveness of the proposed concept of target domain prioritization, as well as the related scheme. In addition, in the experiments concerning the ablation of edge features, the effect of multidimensional edge features is also significantly better than that of the

Table 2. Results of ablation experiments

model	→ Market-1501		→ DukeMTMC	
	Rank-1	mAP	Rank-1	mAP
Baseline	77.7	54.8	62.1	40.7
+NTDP (basic)	84.3	58.1	67.2	44.3
+TDP (basic)	88.6	69.2	73.3	52.8
+TDP (high)	92.7	80.1	80.9	66.3

basic 0/1 edge structure, which shows that the correlation between nodes through multi-dimensional edge features can realize the optimization of the basic message passing mechanism and improve the model performance.

4.4 Hyperparametric Experiments

The DFTM method involves two key hyperparameters: the update parameter of the exponential sliding averageα and the length of the multidimensional edge featuresP. For these two hyperparameters, this subsection conducts experiments on the migration task → Market-1501 to conduct experiments and analyze the results.

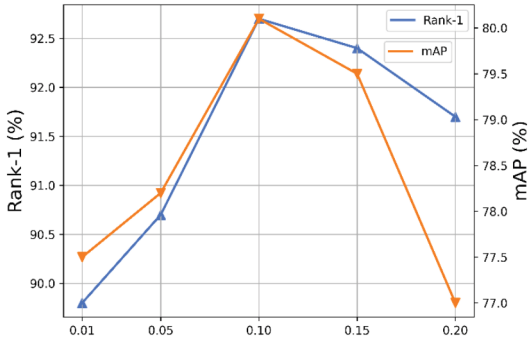

Fig. 3. Experimental comparison of hyperparameter α taking values

The hyperparameter α, which serves as the update parameter for exponential sliding averages, is empirically set to generally be 0.1 or 0.01. In view of this, experimental comparisons are carried out with α values of 0.2, 0.15, 0.1, 0.05, and 0.01, and the results are shown in Fig. 3.

The results show that the hyperparameter α achieves the best performance when set to 0.1. Since α reflects the strategy of updating the center node, setting α to a value that is too large will result in fast updates that are affected by the current batch of samples, leading to instability. On the other hand, if α is set too small, the update speed will be too slow, resulting in worse convergence speed and convergence effect of the network. Considering these factors, the final hyperparameter α is set to 0.1.

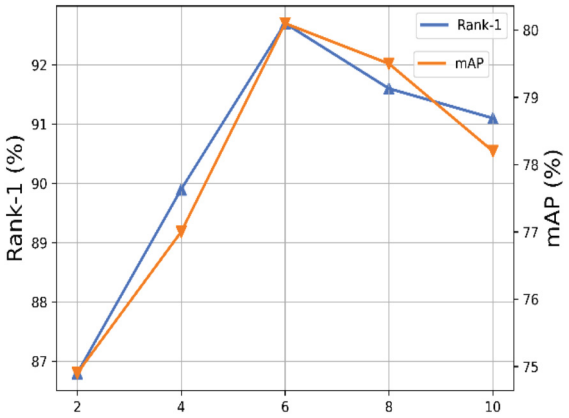

Fig. 4. Experimental comparison of hyperparameter P values

For the hyperparameters P, due to the small number of nodes in the graph convolutional network constructed in the DFTM method, edge features with excessively high dimensions are prone to overfitting. This means that the correlation information between the limited number of nodes is not complex enough to support the learning of edge features with extremely high dimensions. Considering this, the value of P is limited to the range of 2 to 8, and we take values of 2, 4, 6, 8, and 10. The experimental results are shown in Fig. 4, which indicates that the hyperparameter $P = 6$ yields the best results. Additionally, there is no significant difference between the results when using $P = 6$, 8, or 10. However, when using $P = 2$ or $P = 4$, the performance exhibits a more noticeable decline. This indicates that multi-dimensional edge features struggle to fully contribute to lower-dimensional edge features, and increasing the dimensionality is necessary to effectively mine the correlation information between the nodes.

4.5 Visualization of Experimental Results

Domain fusion aims to address the distributional differences between different domains, and its effectiveness can be evaluated using metrics such as Rank-1 and mAP, as well as through visualizing feature distributions using downscaled representations. Specifically, the fused features can be downscaled and projected into a two-dimensional space using t-SNE, enabling visualization of feature fusion and its degree across different domains.

In this subsection, we visualize the features after DFTM processing using t-SNE, as shown in Fig. 5. Different colors represent features from various domains, with gray indicating features from the target domain. In particular, the figures (a), (b), and (c) illustrate scenarios where DFTM is not used, DFTM is used without target domain prioritization, and DFTM is fully employed with target domain prioritization, respectively. Without using DFTM, as shown in (a), the distribution of features from different domains appears more discrete, indicating minimal fusion between domains. The distribution difference between different source domains is small, while the difference with the target domain is significant, leading to limited fusion.When DFTM is used without target domain prioritization, as shown in (b), a certain degree of domain fusion is achieved. However, the

target domains still maintain some distance compared to the source domains.In contrast, in Fig. (c), a superior domain fusion is achieved due to the use of the target domain prioritization mechanism.

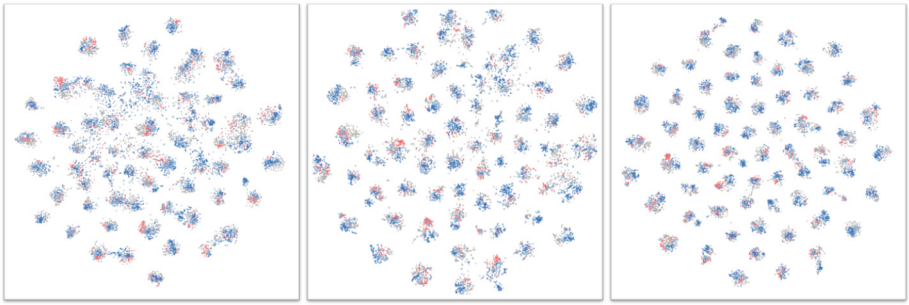

Fig. 5. t-sne visualization

5 Conclusion

In this paper, we propose an unsupervised multi-source domain-adapted pedestrian re-recognition method that combines Domain Fusion based on Target Domain Prioritization and Multidimensional Edge Features (DFTM). The method achieves step-by-step domain fusion through a two-stage graph convolution process. To address the problem of significant distributional differences in the target domain compared to the source domain, we introduce a target domain prioritization mechanism. Additionally, we tackle the challenge of insufficiently mining the deep correlation between nodes by incorporating multidimensional edge features. Our approach is supported by multiple experimental results using four datasets across three migration tasks, demonstrating the superior performance of DFTM and providing strong evidence of its effectiveness.

Acknowledgments. This research received partial support from the R&D plan project in the key scientific research platform of universities in Guangdong Province (No. 2022KSYS016), the Guangdong Province Key Laboratory of Intelligent Detection in Complex Environment of Aerospace, Land and Sea (2022KSYS016), and the Research Platforms and Projects in Ordinary Universities in Guangdong Province (Natural Science) 2019KTSCX216.

References

1. Sun, Y., Zheng, L., Yang, Y., Tian, Qi., Wang, S.: Beyond part models: person retrieval with refined part pooling (and a strong convolutional baseline). In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) Computer Vision – ECCV 2018: 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part IV, pp. 501–518. Springer International Publishing, Cham (2018). https://doi.org/10.1007/978-3-030-01225-0_30
2. Wang, G., Yuan, Y., Chen, X., Li, J., Zhou, X.: Learning discriminative features with multiple granularities for person re identification. In: 2018 ACM Multimedia Conference on Multimedia Conference, MM 2018, Seoul, Republic of Korea, 22–26 October 2018, pp. 274–282 (2018)
3. Zhang, Z., Lan, C., Zeng, W., Jin, X., Chen, Z.: Relation-aware global attention for person reidentification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2020)
4. Song, L., et al.: Unsupervised domain adaptive re-identification: theory and practice. Pattern Recogn. **102**, 107173 (2020)
5. Yang, F., et al.: Asymmetric co-teaching for unsupervised cross-domain person reidentification. In: AAAI, pp. 12597–12604 (2020)
6. Ge, Y., Chen, D., Li, H.: Mutual meanteaching: pseudo label refinery for unsupervised domain adaptation on person re-identification. In: International Conference on Learning Representations (2020)
7. Bai, Z., Wang, Z., Wang, J., et al.: Unsupervised multi-source domain adaptation for person reidentification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 12914–12923 (2021)
8. Yuan, J., et al.: Self-supervised graph neural network for multi-source domain adaptation (2022)
9. Deng, W., Zheng, L., Ye, Q., et al.: Similarity-preserving image-image domain adaptation for person re-identification. arXiv preprint arXiv:1811.10551 (2018)
10. Wei, L., Zhang, S., Gao, W., et al.: Person transfer gan to bridge domain gap for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 79–88 (2018)
11. Zhai, Y., Lu, S., Ye, Q., et al.: Ad-cluster: augmented discriminative clustering for domain adaptive person re-identification. In: Proceedings of the IEEE/ CVF Conference on Computer Vision and Pattern Recognition, pp. 9021–9030 (2020)
12. Mekhazni, D., Bhuiyan, A., Ekladious, G., Granger, E.: Unsupervised domain adaptation in the dissimilarity space for person re-identification. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12372, pp. 159–174. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58583-9_10
13. Huang, Y., Peng, P., Jin, Y., et al.: Domain adaptive attention learning for unsupervised person re-identification. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, no. 07, pp. 11069–11076 (2020)
14. Yang, Q., Yu, H.X., Wu, A., et al.: Patch-based discriminative feature learning for unsupervised person re-identification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3633–3642 (2019)
15. Tao, X., Kong, J., Jiang, M., et al.: Unsupervised domain adaptation by multi-loss gap minimization learning for person re-identification. IEEE Trans. Circuits Syst. Video Technol. **32**(7), 4404–4416 (2021)
16. Qi, L., Liu, J., Wang, L., et al.: Unsupervised generalizable multi-source person reidentification: a domain-specific adaptive framework. Pattern Recogn. **140**, 109546 (2023)

17. Wang, H., Hu, J., Zhang, G.: Multi-source transfer network for cross domain person re-identification. IEEE Access **8**, 83265–83275 (2020)
18. Zheng, L., Shen, L., Tian, L., et al.: Scalable person re-identification: a benchmark. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1116–1124 (2015)
19. Ristani, E., Solera, F., Zou, R., Cucchiara, R., Tomasi, C.: Performance measures and a data set for multi-target, multi-camera tracking. In: Hua, G., Jégou, H. (eds.) ECCV 2016. LNCS, vol. 9914, pp. 17–35. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-48881-3_2
20. Li, W., Zhao, R., Xiao, T., et al.: Deepreid: deep filter pairing neural network for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 152–159 (2014)
21. Xiang, S., Fu, Y., You, G., et al.: Unsupervised domain adaptation through synthesis for person re-identification. In: 2020 IEEE International Conference on Multimedia and Expo (ICME), pp. 1–6. IEEE (2020)
22. Zheng, L., Yang, Y., Hauptmann, A.G.: Person re-identification: past, present and future. arXiv preprint arXiv:1610.02984 (2016)
23. Ye, M., Shen, J., Lin, G., et al.: Deep learning for person re-identification: a survey and outlook. IEEE Trans. Pattern Anal. Mach. Intell. **44**(6), 2872–2893 (2021)
24. Bruna, J., Zaremba, W., Szlam, A., et al. Spectral networks and locally connected networks on graphs. arXiv preprint arXiv:1312.6203 (2013)
25. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907 (2016)
26. Defferrard, M., Bresson, X., Vandergheynst, P.: Convolutional neural networks on graphs with fast localized spectral filtering. Adv. Neural Inf. Process. Syst. **29** (2016)
27. Atwood, J., Towsley, D.: Diffusion-convolutional neural networks. Adv. Neural Inf. Process. Syst. **29** (2016)
28. Hamilton, W., Ying, Z., Leskovec, J.: Inductive representation learning on large graphs. Adv. Neural Inf. Process. Syst. **30** (2017)
29. Velickovic, P., Cucurull, G., Casanova, A., et al.: Graph attention networks. Stat **1050**(20), 10.48550 (2017)
30. Xu, K., Hu, W., Leskovec, J., et al.: How powerful are graph neural networks?. arXiv preprint arXiv:1810.00826 (2018)
31. Gong, L., Cheng, Q.: Exploiting edge features for graph neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9211–9219 (2019)