IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Unsupervised Horizontal Pyramid Similarity Learning for Cross-domain Adaptive Person Re-identification

WENHUI DONG[1,2], PEISHU QU[1], CHUNSHENG LIU[2], YANKE TANG[1], AND NING GAI[1]

[1]College of Physics and Electronic Information, Dezhou University, Dezhou,253023 China (e-mail: dongwh_81@163.com /qupsh@163.com /tyk450@163.com /ning.gai@hotmail.com)

[2]School of Control Science and Engineering, Shandong University, Ji'nan ,250061 China (e-mail: liuchunsheng@sdu.edu.cn)

Corresponding author: Wenhui Dong (e-mail: dongwh_81@ 163.com) and Chunsheng Liu (e-mail: liuchunsheng@sdu.edu.cn).

This work was supported by the National Natural Science Foundation of China (grant Nos. U1931106 and 11673005), Shandong Provincial Natural Science Foundation of China (grant No. ZR2019YQ03), Project of Shandong QingChuang science and technology plan (grant No. 2019KJJ006) and Research Fund for talents of Dezhou University (grant No. 2016KJRC04).

ABSTRACT Although person re-identification has made great progress, unsupervised cross-domain adaptive person re-identification is still a challenging problem. With no labeled data in target domain, the performance may have a significant drop. In this paper, we propose an unsupervised cross-domain adaptive person re-identification framework based on horizontal pyramid similarity learning (UHPS). Firstly, horizontal pyramid features are extracted by dividing the deep feature maps into different number of partial feature bins. These feature bins with diverse scales can incorporate not only the global information but also local information in different spatial scales, making the framework more robust in complex environment. Then, horizontal pyramid similarity learning is proposed with the mechanism of fusing together the internal similarity of the target domain and the similarity between the source domain and target domain. Finally, the unsupervised clustering algorithm DBSCAN embeded with the horizontal pyramid similarity is employed to select training data in the target domain and estimate the pseudo labels in each training iteration, for the purpose of adapting the framework to the target domain. The results on Market1501 and DukeMTMC-reID confirm that the proposed framework can adapt to the target domain effectively and outperforms the state-of-the-art unsupervised cross domain person re-identification approaches.

INDEX TERMS Person re-identification, Unsupervised deep learning, Unsupervised cross domain adaption, Horizontal pyramid similarity learning

I. INTRODUCTION

PERSON re-identification has been widely studied as a specific person retrieval problem, which aims at matching images of a query with images of the same identity across different cameras. Due to the urgent demand of public safety and large scale surveillance, it has drawn lots of attention from academia and industry in recent years. In the person re-identification process, the same person observed in different camera views undergoes significant variations of resolutions, lightings, poses and viewpoints. Because the size of the objects captured by surveillance cameras are often small, a lot of visual details such as facial components are indistinguishable in images, and some of them look similar in appearance. As the number of objects to be distinguished increases, the ambiguities also may increase. These difficulties make person re-identification a very challenging problem.

Recently, many supervised deep learning person re-identification methods have made impressive progress [1]–[6]. In order to obtain robust features, some supervised part-based person re-identification models which focus on learning partial features are poposed. These models are robust to the unavoidable challenges such as occlusion and partial variations. Some approaches also consider integrating global and partial features together to improve the robustness. However, these methods need a large amount of labeled data, which is costly and not suitable for practical application.

Many works focus on unsupervised cross domain person re-identification [7]–[16]. These methods transfer and generalize the learned model trained in source domain with labeled data to the target domain with unlabeled data. Since the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3093083, IEEE Access

surveillance environments are different, domain gap exists between the source domain and target domain. The key problem is how to effectively adapt the model learned in the source domain to the target domain. Some methods focus on learning view-invariant information of person appearance and adapting it to the target domain by attribute and identity alignment [7], [10]. Other methods apply generative adversarial to transfer the style in the image level [11], [13]. The camera-aware similarity inconsistency problem is also considered in [16]. Although these approaches have achieved promising progress, discriminative feature representation and effective domain adaption strategy are still two open problems.

In this paper, we propose an unsupervised cross-domain adaptive person re-identification framework based on horizontal pyramid similarity learning. Although similar work has been done in the supervised deep learning framework [17], [18], it is the first time that exploring the information of pyramid feature bins of different spatial scales in an unsupervised deep learning framework for person re-identification. The framework firstly extracts horizontal pyramid features of the unlabeled images in the target domain, which contain discriminative information in diverse spatial scales. Then, horizontal pyramid similarity learning is proposed with the mechanism of fusing the internal similarity of the target domain and the similarity between the source domain and target domain together. Finally, the unsupervised clustering algorithm DBSCAN embeded with the horizontal pyramid similarity is employed to select training data and estimate the pseudo labels in each training iteration for domain adaption.

We summarize our contributions as follows:

(1)We propose a robust horizontal pyramid feature representation in an unsupervised manner for person re-identification. Horizontal pyramid feature is a collection of feature bins with diverse scales, which can incorporate not only the global information but also local information in different spatial scales. These different feature bins make the model significantly robust in a complex environment.

(2)Horizontal pyramid similarity learning is proposed with the mechanism of fusing together the internal similarity of the target domain and the similarity between the source domain and target domain. Based on the horizontal pyramid similarity learning, training data are selected and the pseudo labels are estimated in each training iteration, for the purpose of adapting the framework to the target domain.

(3)Extensive experiments are conducted on several popular benchmarks including Market-1501 [19] and DukeMTMC-ReID [20], [21] to demonstrate the effectiveness of the proposed method.

The remainder of this paper is organized as follows. We review the related works in Section II. In Section III, we present our proposed method in detail. In Section IV, our proposed algorithm is evaluated by two public large datasets containing images in different surveillance environment. Experimental results and comprehensive analysis are also included in this section. Finally, we conclude this paper.

II. RELATED WORKS

The method proposed in this paper is related to unsupervised person re-identification, partial feature representation and unsupervised cross domain adaption for person re-identification. So we briefly discuss recent research of the three aspects in this section.

A. UNSUPERVISED PERSON RE-IDENTIFICATION

Unsupervised person re-identification aims at exploring discriminative information from unlabeled person images without expensive data annotation, which is more suitable for real applications. Benefit from the success of deep learning, deeply unsupervised person re-identification methods are popular in recent years [22]. some works focus on purely unsupervised person re-identification without any external dataset or identity annoation. Softened similartity learning is proposed in [23] for unsupervised person re-identification. The framework mines the similarity between unlabeled images as a soft constraint and is trained with the softened label distribution. The work in [24] formulates person re-identification as a multi-label classification task. With the proposed memory-based multi-label classifiction loss (MMCL), the framework predicts multi-class labels to effectively identify images of the same identity. The authors also propose the memory-based positive label prediction (MPLP) to improve the accuracy of of label prediction.

Some works use iterative clustering and classification of unlabeled data for person re-identification. Lin et al. [25] propose a bottom-up clustering method to balance the model between the diversity and similarity. To further improve the performance, hierarchical clustering combined with hard-batch triplet loss (HCT) is proposed in [26]. Hierarchical clustering can help explore similarity among samples and hard-batch triplet loss can reduce the influence of hard samples. The attention-driven two-stage clustering (ADTC) [27] uses a voxel attention mechnasim to highlight the feature of images and obtain spatial information. A two stage clustering strategy is also proposed to generate pseudo labels of unlabeled data, which not only can improve the clustering quality but also stabilize the progressive training. Augmented discriminative clustering (AD-cluster) [28] is proposed to estimate and augment person clusters. By alternating density-based clustering and sample generation, AD-cluster aggregates the discrimination ability of the person re-identification model. In order to depress label noise caused by unsupervised clustering, co-teaching technique is employed in asymmetric co-teaching framework (ACT) [29] and noise resistible mutual training (NRMT) [30]. In ACT, the unlabeled samples are divided into inliers and outliers, and the hard samples are selected at the early stage of domain adaption. The NRMT maintains two networks to perform collaborative clustering and mutual instance selection. The collaborative clustering can ease the fitting to noisy samples. The mutual instance selection can help select reliable and informative samples to train the model. By combining the two parts, the performance of the model is improved greatly.

In this paper, we fuse the pyramid feature similarity learning into an unsupervised deep learning framework for person re-identification. Although similar work has been done in the supervised deep learning framework [17], [18], it is the first time that exploring the information of pyramid feature bins of different spatial scales in an unsupervised deep learning framework for person re-identification. So we mainly discuss the performance improvement caused by fusing the pyramid similarity learning into an unsupervised deep learning framework and other strategies are not discussed in this paper. Note strategies such as co-teaching in [29] and [30] can also be combined to our method and better performance will be obtained.

B. PARTIAL FEATURE REPRESENTATION FOR PERSON RE-IDENTIFICATION

Most existing person re-identification methods either explore partial features or consider global features. Global features represent the whole body of the human [31] and are discriminative when the human body can be accurately located. When the person images suffer from heavy occlusions, partial variations or large background clutter, partial features usually achieve better performance by mining discriminative features of body regions [32]. Due to its advantage in handling these unavoidable challenges, lots of works focus on learning partial discriminative feature representations. Some published works employ pose estimation and landmark detection as tools to parse the body and learn the partial features [33], [34]. Some approaches embed the attention mechanism in the deep network to let the model itself decide where to focus [5], [35]. Recently, the pre-defined patches are proposed in some methods [17], [36], which are simple but effective. Sun et al. [36] proposed a part-based convolutional baseline (PCB) with an uniform partition strategy. A refined part pooling (RPP) method is also used to enhance the within-part consistency. Fu et al. [17] divide the deep feature maps horizontally into multiple spatial bins using various pyramid scales. Then, both global feature and partial feature are employed in a supervised deep network to perform person re-identification independently. Zheng et.al [18] propose a corarse-to-fine pyramid model not only can incorporate local and global information,but also can integrate the gradual cues between them. While having achieved promising results, the above supervised methods are costly and time-consuming, because of needing to label sufficient data.

Currently, some unsupervised person re-identfication methods try to learn partial features in the unsupervised deep learning framework [37], [38]. A patch-based unsupervised learning framework [37] is developed for person re-identification. PatchNet is designed to select patches from feature map and learn discriminative features with an unsupervised patch-based discrimminative feature learning loss. An image-level feature learning loss is also proposed to serve as an image level guidance. A self-similarity grouping appoach (SSG) [38] is proposed to exploit the potential similarity from the global body and two local parts in an unsupervised manner. It iteratively conducts grouping and model training in a self-learning manner. Based on SSG, the authors also introduce a semi-supervised person re-identification method named SSG^{+}. Although intergrating the local and global features improves the performance, partial feature bins are divided with a fix scale in SSG, which may miss the gradual infromation between them and limits the capacities of further improving the performance.

A robust horizontal pyramid feature representation in an unsupervised manner is proposed in our work. In fact, horizontal pyramid feature is a collection of feature bins with diverse scales, which can incorporate not only the global information but also local information in different spatial scales. Although the methods proposed in [17] and [18] also use multiple feature bins with different scales, they embed it in a supervised deep learning framework. The deep learning framework, mechanism and the training mode of our approach are quite different from them.

C. UNSUPERVISED CROSS-DOMAIN ADAPTIVE PERSON RE-IDENTIFICATION

Since it is costly to label the dataset of interest, unsupervised cross-domain adaptive person re-identification becomes one popular solution for person re-identification. It utilizes a fully labeled source domain dataset to extract useful information and transfers the information to the target domain dataset of interest. Since the identity labels are not collected in the target domain during training, it is typically viewed as an unsupervised learning task. The unsupervised cross-domain adaptive person re-identification is closely related to unsupervised domain adaption(UDA), which usually learns a common mapping between source and target distributions for the domain invariant representations [39], [40]. However, UDA approaches assume that the two domains have same class labels, while the person identities of source and target datasets in person re-identification are entirely different. Hence, unsupervised domain adaptation approaches cannot be directly utilized for person re-identification.

Many person re-identification works based on domain adaption are proposed [7]–[16], [41]–[45] in recent years. Reducing data distribution discrepancy between two domains and generating discriminative information for target domain are the two topics these works addressing. Recently, it is a popular approach that using GAN generation to transform source domain images into the style of target domain images. With the generated images, the supervised model trained in the souce domain can learn in the unlabeled target domain. The approach in [8] tries to seek camera invariance and domain connectedness by considering the camera-style variations in the generative adversarial network. Wei et al. [11] design a person transfer GAN with constraints to bridge the domain gap. Deng et al. [13] propose a similarity preserving generative adversarial network to preserve the self-similarity of an image and domain dissimilarity of the translated source image and target image. Wang et al. [45] focus on the robustness of the current best person re-ID models and design

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3093083, IEEE Access

a multi-stage network architecture to extract general and transferable features for adversarial perturbations. Although inspiring performance have been achieved, the scalability and stability of the image generation for large scale changing environment are still challenging.

There are also methods that directly mine discriminative information on the unlabeled target dataset with a trained model from source dataset. Wang et al. [7] take advantage of the attribute labels in the source data to transfer knowledge and propose an unsupervised approach that learns the attribute-semantic and identity discriminative features in the target domain. Fan et al. [42] propose a progressive unsupervised learning method to transfer the pretrained model to unseen dataset. It iterates between pedestrian clustering and fine-tuning of CNN, and adds a selection operation between them to improve the original model. LV et al. [43] transfer the spatio-temporal patterns of the source domain to the target domain based on an unsupervised incremental learning algorithm. The camera view information is also utilized in [46]. The authors develop camera-aware domain adaption to reduce the discrepancy across the camera-level sub-domains and creat discriminative information by exploiting temporal continuity in each camera of target domain. The domain-invariant mapping network (DIMN) [47] with an effective meta-learning based trainning strategy and a memory bank module is proposed to match an arbitrqary number of identities in the target domain. Query-Adaptive Convolution (QAConV) [48] treats image matching as finding local correspondences in deep feature maps to make the matching process interpretable and generalizable. A self-paced contrastive learning framework with hybrid memory is propposed in [49], which can dynamically generate supervisory signals in multi-level for feature representation learning. Multiple expert brainstorming network (MEB-Net) [50] accomplishes domain adaption by brainstorming-based mutual learning among different expert models with different architectures. Although these unsupervised cross-domain adaptive person re-identification approaches have achieved promising progress, the performance still needs to be improved. Discriminative feature representation and effective domain adaption strategy are still two open problems in practical application.

Based on the analysis above, we focus on two aspects to improve the performance of unsupervised cross-domain adaptive person re-identification in this paper. Firstly, a robust horizontal pyramid feature representation in an unsupervised manner is proposed. Global feature and partial feature bins with different scales are joined together to incorporate global information and a variaty of local information. Secondly, horizontal pyramid similarity is explored by fusing the internal similarity of the target domain and the similarity between the source domain and target domain together. Based on the pyramid similarity learning, training data are selected and the pseudo labels are estimated in each training iteration to adapt the framework to the target domain.

III. PROPOSED METHOD

The whole framework of the proposed unsupervised cross-domain adaptive person re-identification is showed in Figure 1. The CNN is initially pre-trained by labeled source dataset. Then the learned model extracts horizontal pyramid features of the unlabeled images in the target domain and explores discriminative information in different spatial scales by the unsupervised horizontal pyramid similarity learning. Finally, the unsupervised clustering algorithm DBSCAN embeded with the learned similarity is employed to select training data and estimate the pseudo labels in each training iteration for domain adaption.

A. SOURCE DOMAIN PRE-TRAINED CNN NETWORK

Suppose the source dataset $\{X_s, Y_s\}$ in source domain $\mathbb{S}$ has N_s labeled person images. Each image x_s^i in X_s has an assosicated identity label $y_s^i \in Y_s, i = 1, 2, \cdots, N_s$. The identity number of the source dataset is P_s. We modify the structure of ResNet-50 [51] pre-trained on ImageNet [52] as backbone of the deep framework. Figure 2 shows the structure. The last FC layer is replaced by two additional FC layers. The first one (FC1) has 2048 outputs and the last one (FC2) has P_s outputs. When training in the source domain, labeled samples are fed into the baseline network, which use cross entropy loss and triplet loss [53] as the training criterion. The cross entropy loss is employed with FC2 by treating the training as a classification problem, and triplet loss is employed with FC1 by treating the training as a verification problem. The advantage of the pre-trained CNN network is that training process can make full use of the information from classification and verification [38].

B. HORIZONTAL PYRAMID FEATURE EXTRACTION

Horizontal pyramid feature is a collection of feature bins with diverse spatial scales. Each feature bin is obtained by dividing the deep feature map with a specific scale. Specifically, given a specific scale σ, the feature map is horizontally partitioned into 2^σ uniform pieces. Then feature bins can be obtained after executing average pooling on these sub-maps. By repeatly dividing the feature map with different scales in the scale set $\{\sigma|\sigma \leq \sigma_0\}$, feature bins with different spatial scales can be obtained. The total number of feature bins N_{σ_0} for a feature map can be calculated in formula (1).

Figure 3 shows an example of horizontal pyramid feature extraction with the scale set $\{\sigma|\sigma \leq 3\}$. In this case, $\sigma \in \{0, 1, 2, 3\}$ and the feature map will be horizontally partitioned into 2^0, 2^1, 2^2 and 2^3 uniform pieces independedly. The total number of feature bins is 15. When $\sigma = 0$, the feature map is not divided and the global feature is obtained. With different scales, feature bins with different sptial scales can be obtained in the same image. So the collection of all the feature bins can incorporate not only the global information but also local information in different spatial scales, making the framework more robust in complex environment.

We assume an unlabeled dataset X_T in target domain $\mathbb{T}$ has N_T images. Horizontal pyramid features are ex-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3093083, IEEE Access

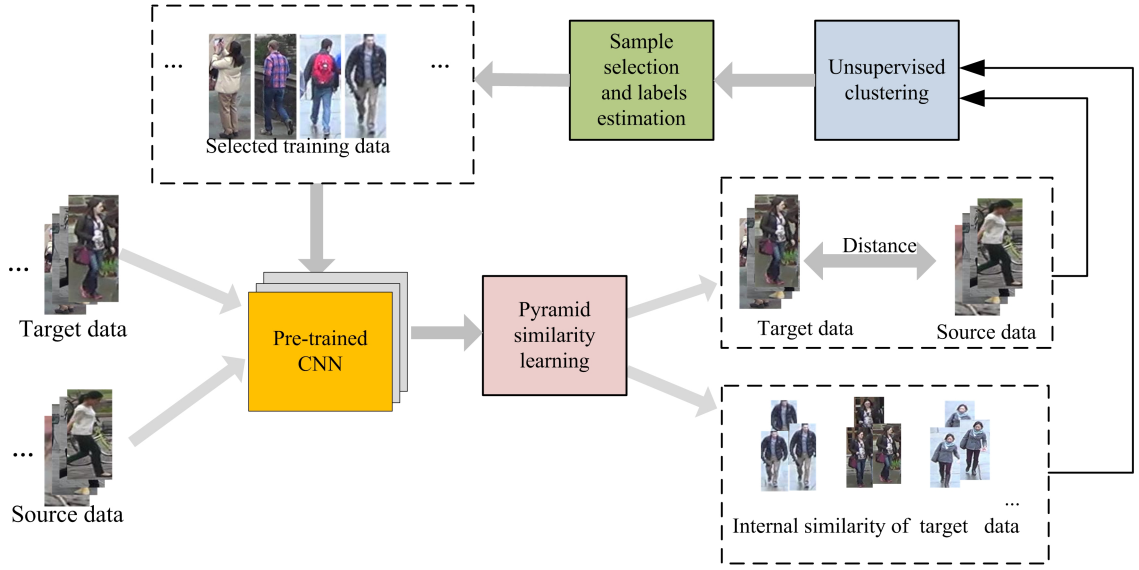

FIGURE 1. The framework of the proposed method: The pre-trained CNN model extracts horizontal pyramid features of the unlabeled images in the target domain. The internal similarity of the target domain and the similarity between the source domain and target domain are fused together for horizontal pyramid similarity learning. The unsupervised clustering algorithm selects training samples and estimates pseudo labels. The whole framework will be iteratively trained.

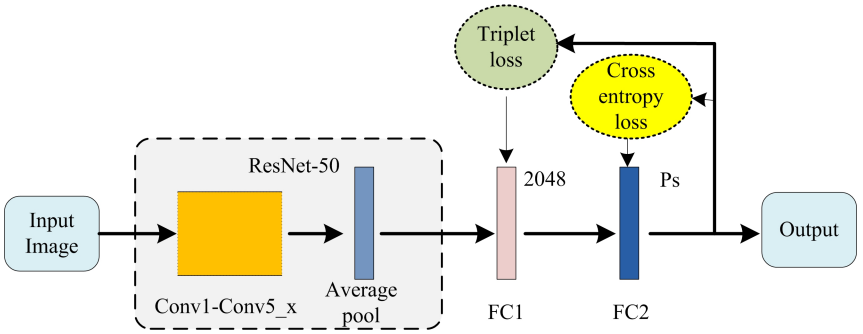

FIGURE 2. The structure of the pre-trained CNN network: The framework is a variant of ResNet-50,the last FC layer of which is replaced with two additional FC layers. Cross entropy loss and triplet loss are used as training criterion.

tracted repeatedly of all the unlabeled images $x_T^j \in X_T, j = 1, 2, \cdots, N_T$ and different features $\mathbf{f}_{T,k}^j, k = 1, 2, \cdots, 2^\sigma, j = 1, 2, \cdots, N_T$ can be obtained with different scales for each image. All the following feature vectors set in formula (2) are fed into the similarity learning module.

$$N_{\sigma_0} = \sum_{\sigma=0}^{\sigma_0} 2^\sigma \tag{1}$$

$$\begin{cases} \mathbf{f}_{T,2^0} = \{\mathbf{f}_{T,1}^1, \mathbf{f}_{T,1}^2, \cdots, \mathbf{f}_{T,1}^{N_T}\} \\ \mathbf{f}_{T,2^1} = \{\{\mathbf{f}_{T,1}^1, \mathbf{f}_{T,2}^1\}, \{\mathbf{f}_{T,1}^2, \mathbf{f}_{T,2}^2\}, \cdots, \{\mathbf{f}_{T,1}^{N_T}, \mathbf{f}_{T,2}^{N_T}\}\} \\ \quad \vdots \\ \mathbf{f}_{T,2^{\sigma_0}} = \{\{\mathbf{f}_{T,1}^1, \mathbf{f}_{T,2}^1, \cdots, \mathbf{f}_{T,2^{\sigma_0}}^1\}, \{\mathbf{f}_{T,1}^2, \mathbf{f}_{T,2}^2, \cdots \mathbf{f}_{T,2^{\sigma_0}}^2\}, \\ \cdots, \{\mathbf{f}_{T,1}^{N_T}, \mathbf{f}_{T,2}^{N_T}, \cdots, \mathbf{f}_{T,2^{\sigma_0}}^{N_T}\}\} \end{cases} \tag{2}$$

C. HORIZONTAL PYRAMID SIMILARITY LEARNING

In this section, we propose the horizontal pyrammid similarity learning to explore discriminative information in different spatial scales. Horizontal pyramid feature bins with a same spatial scale are put together to learn the similarity of a specific scale. After learning similatity of all the scales independently, we can obtain global information and local information in different spatial scales amnong the samples. So horizontal pyramid similatity can be seen as a collection of similarity learned from horizontal pyramid feature bins with different spatial scales.

Besides, in order to obtain effective domain adaptation, training data in the target domain need to be selected. Because the model is initially trained by the source data, samples similar to the source data can help the model explore correct information of the target domain, especially in the early iteration stages when the framework knows little about the target domain. Besides, learning the similarity among different person images in target domain can also help the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3093083, IEEE Access

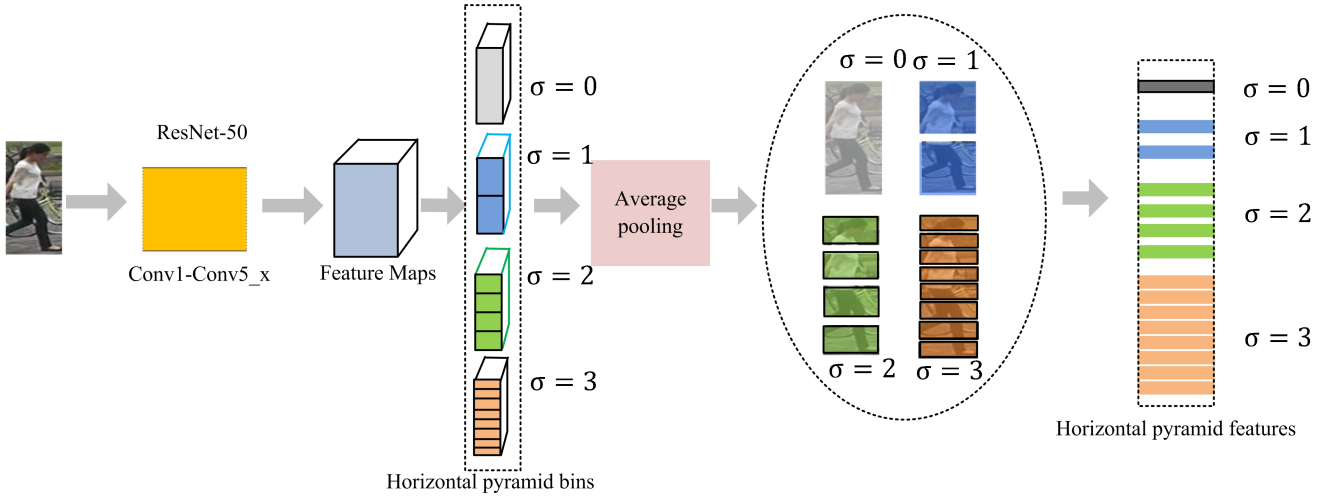

FIGURE 3. An example of the horizontal pyramid feature extraction for $\sigma_0 = 3$: $\sigma \in \{0, 1, 2, 3\}$ and the feature map will be horizontally partitioned into 1,2,4 and 8 uniform pieces. The total number of feature bins is 15.

model discover the correct identification. Hence, the similarity is designed to fuse the the internal similarity of the target doamin and similarity between the source domain and target domain together.

1) Similarity learning between source and target domain

To ensure the effect of domain adaption, we should concern the overlap between the source and target domain. Similar samples of the source data can help the model to explore correct information of the target domain. As mentioned in [54], minimizing the function in formula (3) can help to enhance the similarity.

$$E_{X_T \sim \mathbb{T}}[inf_{X_S \sim \mathbb{S}} \|\mathbf{f}_T - \mathbf{f}_S\|] \tag{3}$$

where, $\mathbf{f}_T$ and $\mathbf{f}_S$ are the features of target and source samples respectively. However, it is hard to get optimal solution due to the infimum. So we approximate the formula by selecting samples with smaller $inf_{X_S \sim \mathbb{S}} \|\mathbf{f}_T - \mathbf{f}_S\|$. Specifically, we adopt the similarity measurement in formula (4) to search the samples in target domain.

$$d_S(\mathbf{f}_T^i) = 1 - e^{(-(\|\mathbf{f}_T^i - N_s(\mathbf{f}_T^i)\|))} \tag{4}$$

where, $N_s(\mathbf{f}_T^i)$ means the nearest neighbor of the $\mathbf{f}_T^i$ in the source domain. $\mathbf{f}_T^i$ reprsents the feature of ith images in X_T. A smaller $d_S(\mathbf{f}_T^i)$ means a higher confidence that the target sample is similar to the source domain.

2) Internal similarity learning within target domain

In order to accurately measure the similarity of samples in the target domain, the contexture of each sample is considered when calculating the distance. We adopt the k-reciprocal nearest neighbors [55] to represent the contexture of the sample in this paper. Each sample is encoded as a k-reciprocal vector, and a variation of Jaccard distance between the vectors is used as the distance metric for similarity measurement. Specifically, for a sample feature pair in target domain $(\mathbf{f}_T^i, \mathbf{f}_T^j), i, j = 1, 2, \ldots, N_T, i \neq j$, we can calculate the distance of the two samples by formula (5).

$$d_v(\mathbf{f}_T^i, \mathbf{f}_T^j) = 1 - \frac{\sum_{k=1}^{N_T} \min(v_{i,k}, v_{j,k})}{\sum_{k=1}^{N_T} \max(v_{i,k}, v_{j,k})} \tag{5}$$

where, $v_{i,k} = e^{-\|\mathbf{f}_T^i - \mathbf{f}_T^k\|^2}$ if $\mathbf{f}_T^k$ is a k-reciprocal nearest neighbor of $\mathbf{f}_T^i$, else $v_{i,k} = 0$. $v_{j,k}$ has the same definition.

3) Similarity fusion

For a specific scale σ, the two kinds of similarity are fused together using formula (6). The similarity of all scales can be calculted independently by formula (6) and the collection of them constructs the pyramid similarity.

$$d(\mathbf{f}_{T,k}^i, \mathbf{f}_{T,k}^j) = (1-\beta) d_v(\mathbf{f}_{T,k}^i, \mathbf{f}_{T,k}^j) + \beta(d_S(\mathbf{f}_{T,k}^i) + d_S(\mathbf{f}_{T,k}^j)) \tag{6}$$

where, $\mathbf{f}_{T,k}^i$ and $\mathbf{f}_{T,k}^j$ are the kth pyramid feature bins of image x_T^i and x_T^j , $k \in 1, 2, \cdots, 2^\sigma$, $\beta \in [0, 1]$ is a balancing parameter.

D. CLUSTERING-GUIDED TRAINING SAMPLE SELECTION AND PSEUDO LABEL ESTIMATION

After learning the pyramid similarity of all data pairs in the target domain by formula (6) of a specific pyramid scales σ, a distance matrix M_σ can be obtained. Then, unsupervised clustering algorithm DBSCAN [56] is applied on the distance matrix M_σ to estimate the pseudo labels of the feature bins. Since only data within the scanning radius can be clustered in DBSCAN, training samples can be selected naturally by setting the scanning radius during the execution of the clustering algorithm. Specifically, all the pair distances calculated by formula (6) are firstly sorted from small to large, then the scan radius ε_σ of DBSCAN is set as the mean value of top pN distance of data pairs in scale σ. Where p is the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3093083, IEEE Access

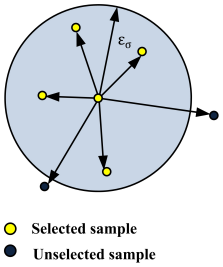

FIGURE 4. The demonstration of sample selection: Distance of each data pair in target domain is calculated by (6). Sample pairs whose distance is within the scan radius ε_σ will be selected as training samples.

percentage and N is the number of possible data pairs in the target domain. Figure 4 demonstrates the sample selection strategy. Only sample pairs whose distance is within the scan radius ε_σ will be selected as the training samples. Since the clustering algorithm is applied in all scales of horizontal pyramid feature sets, each person image will obtain pseudo labels in different scales. All the sample features in different scales are used to train the model. The distance matrix M_σ will be re-calculated for sample pairs re-selection in each training iteration. The selected samples help the model learn more of the target domain, and the learned model can select better traing sample pairs in next training iteration.

E. LOSS FUNCTION

For unsupervised training in target domain, we join the batch-hard triplet loss of different scales and the cross entropy loss of different scales together as framework loss. For a specific scale σ, the triplet loss is formulated as follows:

$$L_{(triplet,\sigma)} = \sum_{k=1}^{2^\sigma}\sum_{i=1}^{P}\sum_{a=1}^{K}[m + \max_{p=1,2,\cdots,K} \|\mathbf{f}_{a,k}^i - \mathbf{f}_{p,k}^i\|_2 - \min_{\substack{j=1,2,\cdots,P \\ n=1,2,\cdots,K, i\neq j}} \|\mathbf{f}_{a,k}^i - \mathbf{f}_{n,k}^j\|_2]_+ \tag{7}$$

where, $\mathbf{f}_{a,k}^i$, $\mathbf{f}_{p,k}^i$, $\mathbf{f}_{n,k}^j$ are the kth feature bins extracted from the anchor, positive and negative samples respectively, m is the margin hyper-parameter. Each mini-batch is constructed by randomly sampling P identities and K instances of feature bins in the same scale.

The cross entropy loss of the specific scale σ is also employed and can be formulated as follows:

$$L_{(ce,\sigma)} = \sum_{k=1}^{2^\sigma}\sum_{i=1}^{P}\sum_{a=1}^{K} l_{ce}(y_{a,k}^i, \hat{y}_{a,k}^i) \tag{8}$$

where, $y_{a,k}^i$ and $\hat{y}_{a,k}^i$ are the pseudo identity and the prediction of the samples respectively, l_{ce} is the cross entropy loss.

The final loss function is the sum of all items of all the scales:

$$L_{target} = \sum_{\sigma} L_{(triplet,\sigma)} + L_{(ce,\sigma)} \tag{9}$$

F. THE ALGORITHM OF WHOLE FRAMEWORK

Now the whole framework can be concluded in Algorithm 1. At the beginning, the baseline model is trained in the source domain and is called as $\mathbf{f}^0$. Then horizontal pyramid similarity learning and DBSCAN clustering are fused together for training data selection and pseudo label estimation. The model will be updated iteratively.

Algorithm 1 Unsupervised horizontal pyramid similarity for cross-domain adaptive person re-identification

Input: source domain dataset S, unlabeled target domain dataset T with N_T samples, the pyramid scale parameter σ_0, the minimum size of a cluster N_1, the iteration number N_2

Output: the model $\mathbf{f}$ for target domain

1: Train the model $\mathbf{f}^0$ on the source domain dataset S;
2: **For** σ in $\{\sigma|\sigma \leq \sigma_0\}$**do**
3: Obtain pyramid features in S and T: $S_\sigma^0 = \mathbf{f}^0(S,\sigma)$, $T_\sigma^0 = \mathbf{f}^0(T,\sigma)$;
4: Compute the distance matrix M_σ^0 on S_σ^0 and T_σ^0 by formula (6);
5: Compute the scan radius ε_σ of DBSCAN clustering ;
6: Estimate labels for $\{x_T^i, i \in \{1,2,\cdots,N_T\}\}$: $(y_{T,1}^i, y_{T,2}^i, \cdots, y_{T,2^\sigma}^i) = DBSCAN(M_\sigma^0, \varepsilon_\sigma, N_1)$;
7: **EndFor**
8: Construct the training data: $D^0 = \{x_T^i : (y_{T,1}^i, y_{T,2}^i, \cdots, y_{T,2^\sigma}^i), \sigma \in \{\sigma \leq \sigma_0\}\}$
9: Train the model by D^0 and obtain $\mathbf{f}^1$
10: **For** $j = 1$ to N_2
11: Repeat step 2 to step 9 ;
12: Obtain $\mathbf{f}^j$ by training with dataset D^j;
13: **EndFor**

IV. EXPERIMENT

In this section, we evaluate the proposed method on two large person re-identification benchmark datasets Market1501 [19] and DukeMTMC-reID [21]. Ablation study is applied to evaluate the key components and parameters in the proposed method. We also compare our method with the state-of-the-art unsupervised person re-identification methods.

A. DATASETS AND EXPERIMENT SETTINGS

Market1501 consists of 32,668 images of 1,501 identities captured by six cameras in an open environment. The training set has 751 identities with 12,936 images and the test set contains 19,732 images of 750 identities. DukeMTMC-reID is also a challenging person re-identification dataset. It has 16,522 training images, 2,228 query images, and 17,661 gallery images, totally containing 1,812 identities in 8 camera views. The two datasets are challenging for person re-identification and undergo significant variations of resolutions, lightings, poses, occlusion and viewpoints. In our experiment, we follow the standard train/test split of the two datasets. The Cumulative Matching Characteristic (CMC) and mean Average Precision (mAP) are applied as

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3093083, IEEE Access

the performance evaluation metrics following single query setting. The evaluation packages provided by [19]and [21] are used respectively.

B. IMPLEMENTATION DETAILS

The input images are resized to $256 \times 128 \times 3$ and augmented with random cropping, flipping and random erasing. The batch size is 128 (PK sampling with $P = 16$, $K = 8$) and Adam with decay 0.0005 is chosen as optimizer. When training the baseline model on the source dataset, the epoch is set to 120 and the learning rate is changed from 3×10^{-4} to 3×10^{-5} after 100 epochs. When training the model in the unsupervised mode on target dataset, the initial learning rate is set to 3×10^{-4} , the training epoch is 70, $N1 = 4$ and $N2 = 20$. Our algorithm is implemented on Pytorch platform and trained with two NVIDIA TITAN X GPUs. All the experiments discussed follow the same settings.

C. ABLATION STUDY

In order to verify the effectiveness of each component and parameter settings of UHPS, several ablation experiments are designed on Market1501 and DukeMTMC-reID datasets, including different number of pyramid scales, different distance metrics. The different parameters setting such as different values of p and β are also tested in this section. [1]

1) Effectiveness of pyramid structure

We compare the performance of UHPS with different pyramid scales. Table 1 and Table 2 show the results with different pyramid scales on the Market1501 and DukeMTMC-reID. The value of pyramid scale σ_0 determines the number of feature bins. Four cases of pyramid scale $\sigma_0 = 0, 1, 2, 3$ are tested and the results show that $\sigma_0 = 2$ achieves the best. When $\sigma_0 = 0$, only global feature are extracted. From Table 1, we can observe that when the pyramid scale increases from 0 to 2 on Market1501 dataset, the Rank-1 improves from 75.6% to 81.2%, and mAP improves from 53.6% to 59.1%. However, the Rank-1 and mAP drop to 79.6% and 58% when the pyramid scale is set to 3. A similar trend can be obtained on DukeMTMC-reID in Table 2. These results infer that the pyramid structure can improve the performance of the person re-identification framemork by combining the global and local discriminative information. However, too many feature bins may produce redundant information and yield worse results. So we finally adopt $\sigma_0 = 2$ and $\sigma = 0, 1, 2$ in this work.

2) Effectiveness of distance metrics

In this work, we select training samples considering both the similarity between the source domain and target domain and the similarity of the samples in the target domain. To verify the performance of the similarity measure module, different distance metrics are compared in Table 3 and Table 4, where

[1]When one parameter is tested in the experiment,other parameters and settings of the framework are set to the optimal values.

TABLE 1. Performance of UHPS with different pyramid scales on Market1501

Methods	σ_0	σ	Feature bins	Rank-1	Rank-5	Rank-10	mAP
UHPS0	0	0	1	75.6	89.7	93.8	53.6
UHPS1	1	0,1	$1+2$	80.2	90.8	92.4	58.4
UHPS2	2	0,1,2	$1+2+4$	**81.2**	**91.3**	**93.2**	**59.1**
UHPS3	3	0,1,2,3	$1+2+4+8$	79.6	90.1	92	58

TABLE 2. Performance of UHPS with different pyramid scales on DukeMTMC-ReID dataset

Methods	σ_0	σ	Feature bins	Rank-1	Rank-5	Rank-10	mAP
UHPS0	0	0	1	68.3	80.3	83.8	49.2
UHPS1	1	0,1	$1+2$	72.9	80.8	82.9	53
UHPS2	2	0,1,2	$1+2+4$	**73.8**	**81.4**	**84.1**	**54.4**
UHPS3	3	0,1,2,3	$1+2+4+8$	72.1	80.1	82	52.4

d_E represents Euclidean distance, d_v is the Jaccard distance, d_s is the distance between source domain and target domain, and $d_v + d_s$ means our distance metric introduced in section III-C. Taking the results on the Market1501 dataset in Table 3 as example, it achieves 79.3% and 58.1% on the Rank-1 accuracy and mAP when only considering d_v, which are higher than the results of d_E. This infers the effectiveness of d_v. However, our distance metric achieves the highest accuracy on the two datasets showing the advantage of the combinations of d_v and d_s. The similar results can also be seen in Table 4 on DukeMTMC-reID dataset.

3) Parameters analysis

As the analysis in section III-D, parameters β and p are the two important parameters, which directly influence training sample selection and label estimation. In this experiment, we test our method with a series of different β and p respectively on Market-1501. The final Rank-1, Rank-5, Rank-10 and mAP results are shown in Table 5-6. In addition, Figure 5-8 demonstrate the Rank-1 and mAp variation curves in 20 iterations with different β and p. As can be seen from Table

TABLE 3. Comparison of different distance metrics on Market1501

Distance	Rank-1	Rank-5	Rank-10	mAP
d_E	78.6	89.7	91.5	57.4
d_v	79.3	90.2	92.2	58.1
d_v+d_s	**81.2**	**91.3**	**93.2**	**59.1**

TABLE 4. Comparison of different distance metrics on DukeMTMC-ReID

Distance	Rank-1	Rank-5	Rank-10	mAP
d_E	71.6	79.4	83.1	52.6
d_v	72.3	80.2	83.8	53.2
d_v+d_s	**73.8**	**81.4**	**84.1**	**54.4**

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3093083, IEEE Access

TABLE 5. Performance comparison of different value of β on Market1501

β	Rank-1	Rank-5	Rank-10	mAP
0.05	79.7	90.2	92.5	58.6
0.1	**81.2**	**91.3**	**93.2**	**59.1**
0.3	79.3	90	91.2	57.8
0.5	79	89.7	90.6	57.2
0.7	77.9	88.3	90.1	55.9

TABLE 6. Performance comparison of different value of p on Market1501

p	Rank-1	Rank-5	Rank-10	mAP
1.1×10^{-3}	78.7	87.6	89.6	55
1.3×10^{-3}	79.6	89.8	92	56.2
1.5×10^{-3}	81	90.9	92.7	58.6
1.7×10^{-3}	**81.2**	**91.3**	**93.2**	**59.1**
2.0×10^{-3}	80.4	90.4	92.1	58
2.2×10^{-3}	78.8	88.1	91.8	55.6

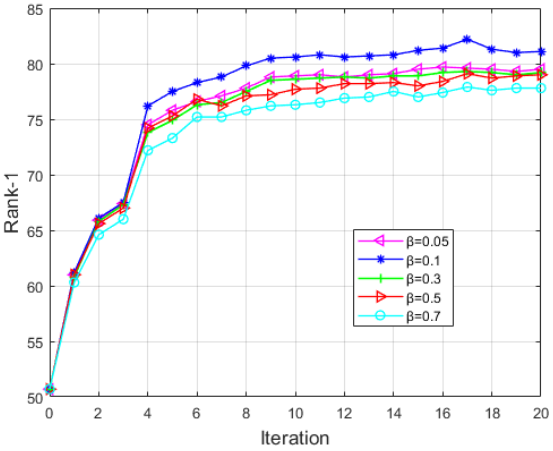

FIGURE 5. Rank-1 curves with different β : it shows the Rank-1 performance curves with different value of β during 20 iterations. $\beta = 0.05, 0.1, 0.3, 0.5, 0.7$ are tested respectively.

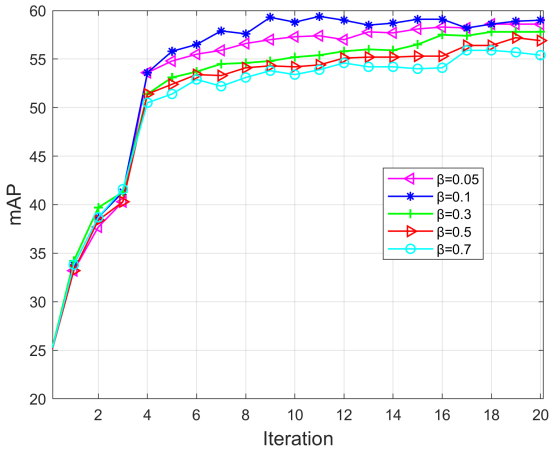

FIGURE 6. mAP curves with different β: the mAP performance curves with different value of β during 20 iterations are showed. $\beta = 0.05, 0.1, 0.3, 0.5, 0.7$ are test respectively.

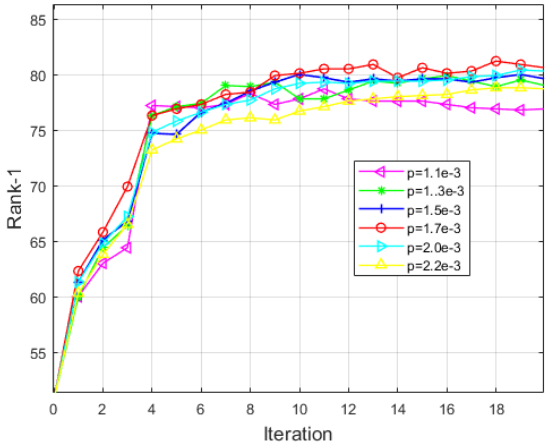

FIGURE 7. Rank-1 curves with different p: the Rank-1 performance with different value of p. $p = 1.1 \times e^{-3}, 1.3 \times e^{-3}, 1.5 \times e^{-3}, 1.7 \times e^{-3}, 2.0 \times e^{-3}, 2.2 \times e^{-3}$ are tested repectively.

5, Figure 5 and 6, performance changes with different parameter β and best performance can be gotten when $\beta = 0.1$. Table 6, Figure 7 and 8 show the performance of different value of p. We change p from 1.1×10^{-3} to 2.2×10^{-3}. Because the dataset is very large, small change of p will have a large impact on the accuracy. In our test, it achieves the best performance when $p = 1.7 \times 10^{-3}$.

D. COMPARISON WITH STATE-OF-THE-ART UNSUPERVISED MODELS

We compare our UHPS with the state-of-the-art unsupervised models for person re-identification on Market1501 and DukeMTMC-reID in Table 7 and Table 8 respectively. The models include SPGAN [13], TJ-AIDL [7], MMFA [58], HHL [8], ARN [12], PDA-Net [15], MAR [57], ENC [14], UDA [54] and SSG [38]. Because we mainly discuss the performance improvement caused by fusing the pyramid similarity learning into an unsupervised deep learning framework and no other strategies are used, the models that having other strategies to improve the performance are not chosen to compare with our proposed method. The most related and newest work is SSG. Note strategies such as co-teaching can also be combined to our method and better performance will be obtained.

It can be observed that the performance drops largely when directly transferring the model trained on the source dataset to target set. Specifically, the supervised baseline trained on DukeMTMC-reID achieves 93.2% in rank-1 accuracy and 80.7% in mAP when tested on DukeMTMC-reID, but it drops to 50.7% and 23.9% when directly transferring to Market1501. A similar drop can be observed when directly trans-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3093083, IEEE Access

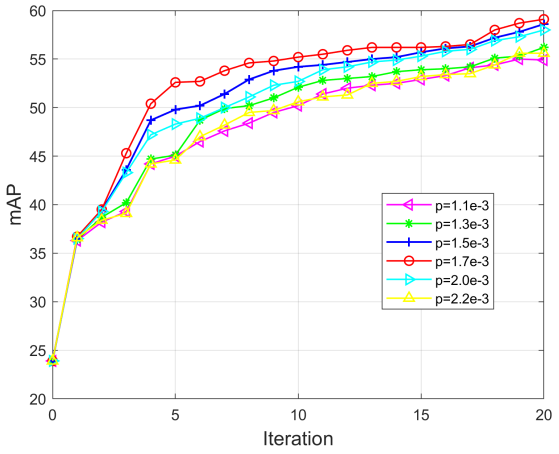

FIGURE 8. mAP curves with different p: the mAP performance curves with different value of p during 20 iterations are showed. In the experiment, $p = 1.1 \times e^{-3}, 1.3 \times e^{-3}, 1.5 \times e^{-3}, 1.7 \times e^{-3}, 2.0 \times e^{-3}, 2.2 \times e^{-3}$ are tested repectively.

ferring the model trained by Market1501 to DukeMTMC-reID. Although all the unsupervised domain adaptation methods shown in Table 7 and Table 8 are proposed to reduce this performance drops, our UHPS outperforms all of them in Rank-1, Rank-5, Rank-10, and mAP.

Comparisons between UHPS and state-of-the-art approaches on Market1501 are shown in Table 7. The results show that our UHPS is the best method and achieves the Rank-1 accuracy of 81.2% and the mAP of 59.1%. The second best method is SSG, which is also based on partial similarity learning. However, it splits the feature maps into the pre-defined two parts. In contrast, our UHPS splits the feature maps according to the pyramid scales, which is more robust. As can be seen in Table 7, our result is $1.2\%$and 0.8% higher in terms of Rank-1 accuracy and mAP when compared to SSG.

Person re-identification results on DukeMTMC-ReID are shown in Table 8. The dataset has eight different cameras and the size of person bounding box varies largely across different camera views. Our UHPS still has the best performance on this challenging dataset and achieves the Rank-1 accuracy of 73.8% and the mAP of 54.4%. Compared with the second best unsupervised method, our result is 0.8% and 1.0% higher on Rank-1 accuracy and mAP respectively.

From the experiments on the two challenging datasets, the effectiveness of the proposed UHPS can be successfully verified. In additon, we also test the stability of our proposed algorithm by runing the algorithm for 5 times on Market1501 dataset, the fluctuation range of rank-1 accuracy and mAP are 80.8 ± 0.9 and 58.8 ± 1 respectivly .

V. CONCLUSION

In this work, we fuse the horizontal pyramid similarity learning into the unsupervised cross-domain adaptive person re-identification framework. The horizontal pyramid segmentation of the feature map can help obtain the discriminative information from coarse to fine and finally form a more robust feature representation. The horizontal pyramid similarity learning that considers both the similarity between the source domain and target domain and the internal similarity of target domain also improves the performance. Extensive ablation studies and comparisons demonstrate the effectiveness of the proposed method. Due to the limitation of hardware resources, the proposed method are not tested on more backbone architectures, which will be done in the future.

TABLE 7. Performance comparisons on Market1501 with the state-of-the-art unsupervised cross-domain adaptive person re-ID methods(Source: DukeMTMC-reID)

Method	Rank-1	Rank-5	Rank-10	mAP
Supervised baseline	93.2	97.9	98.6	80.7
Direct transfer	0.7	7.7	4	3.9
SPGAN [13]	57.7	75.8	82.4	26.7
TJ-AIDL [7]	58.2	74.8	81.1	26.5
MMFA [58]	56.7	75	81.8	27.4
HHL [8]	62.2	78.8	84	31.4
ARN [12]	70.3	80.4	86.3	39.4
MAR [57]	67.7	81.9	-	40
ENC [14]	75.1	87.6	91.6	43
PDA-Net [15]	74.2	86.3	90.2	47.6
UDA [54]	75.8	89.5	93.2	53.7
SSG [38]	80	90	92.4	58.3
UHPS	**81.2**	**91.3**	**93.2**	**59.1**

TABLE 8. Performance comparisons on DukeMTMC-reID with the state-of-the-art unsupervised cross-domain adaptive person re-ID methods(Source: Market1501)

Method	Rank-1	Rank-5	Rank-10	mAP
Supervised baseline	83.2	92.7	94.5	71.2
Direct transfer	0.7	5.3	3	1.9
SPGAN [13]	46.4	62.3	68	26.2
TJ-AIDL [7]	44.3	59.6	65	23
MMFA [58]	45.3	59.8	66.3	24.7
HHL [8]	46.9	61.0	66.7	27.2
ARN [12]	60.2	73.9	79.5	33.4
MAR [57]	67.1	79.8	-	48
ENC [14]	63.3	75.8	80.4	40.4
PDA-Net [15]	63.2	77	82.5	45.1
UDA [54]	68.4	80.1	83.5	49.0
SSG [38]	73	80.6	83.2	53.4
UHPS	**73.8**	**81.4**	**84.1**	**54.4**

REFERENCES

[1] X. B. chang,T. M. Hospedales and T. Xiang, "Multi-level factorisation net for person re-identification," in *Proc. IEEE CVPR,* Salt Lake City, USA, 2018, pp. 2109–2118.

[2] J. R. Yang, W. S. Zheng, Q. Z. Yang, Y. C. Chen and Q. Tian, "Spatial-temporal graph convolutional network for video-based person re-identification," in *Proc. IEEE CVPR,* 2020, pp. 3289–3299.

[3] Y. F. Sun, L. Zheng, Y. Yang, Q. Tian and S. J. Wang "Beyond part models: person retrieval with refinedpart pooling (and a strong convolutional baseline)," in *Proc. IEEE ECCV,* Munich,Germany, 2018, pp. 480–496.

[4] H. Zhao, M. Tian, S. Sun, J. Shao, J. Yan, S. Yi, X. Wang and X. Tang "Spindle net: person re-identification with human body region guided feature decomposition and fusion," in *Proc. IEEE CVPR,* Hawaii, 2017, pp. 1077–1085.

[5] H. Yao, S. Zhang, Y. Zhang, J. Li and Q. Tian. " Deep representation learning with part loss for person re-identification," arXiv preprint arXiv:1707.00798, 2017. [Oneline]. Available: https://arxiv.org/abs/1707.00798.

[6] C. Liu, X. J, Chang and Y.D.Shen, "Unity style transfer for person re-identification," in *Proc. IEEE CVPR,* Vitural,2020, pp. 6887–6896.

[7] J. Y. Wang, X .T. Zhu, S. G. Gong and W. Li, "Transferable joint attribute-identity deep learning for unsu-pervised person re-identification," in *Proc. IEEE CVPR,* Salt Lake City, USA, 2018, pp. 2275–2284.

[8] Z. Zhong, L. Zheng, S. Z. Li and Y. Yang, "Generalizing a person retrieval model hetero-and homogeneously," in *Proc. IEEE ECCV,* Munich,Germany, 2018, pp. 172–188.

[9] Y. P. Zhai, S. J. Lu, Q. X. Ye, X. B. Shan, J. chen, R. R. Ji and Y. H. Tian, 'AD-Cluster:augmented discriminative clustering for domain adaptive person re-identification," in *Proc. IEEE CVPR,* Virtual ,2020, pp.9021–9030.

[10] S. Lin, H. L. Li, C. T. Li, and A. ChichungKot, " Multi-task mid-level feature alignment network for un-supervised cross-dataset person re-identification," in *Proc. BMVC,*, California, USA, 2018, pp. 172–188.

[11] L. H. Wei, S. L. Zhang, W. Gao, and Q. Tian, " Person transfer gan to bridge domain gap for person re-identification,"in *Proc. IEEE CVPR,* Salt Lake City, USA, 2018, pp. 79–88.

[12] G. C. Wang, J. H. Lai, W. Q. Liang and J. R. Wang, "Smoothing adversarial domain Attack and p-memory reconsolidation for cross-domain person re-identification," in *Proc. IEEE CVPR,* 2020, pp.10568–10577.

[13] W. J. Deng, L. Zheng, Q. X. Ye, G. L. Kang, Y. Yang and J. B. Jiao, " Image-image domainad aptation with preserved self-similarity and domain dissimilarity for person re-identification,"in *Proc. IEEE CVPR,* Salt Lake City, USA, 2018, pp. 994–1003.

[14] Z. Zhong, L. Zheng, Z. M. Luo, S. Z. Li and Y. Yang, "Invariance matters: exemplar memory for domain adaptive person re-identification," in *Proc. IEEE CVPR,* California, USA, 2019, pp. 598–607.

[15] Y. J. Li, C. S. Lin and Y. C. F. Wang, "Cross-dataset person re-Identificationvia unsupervised pose disentanglement and adaptation." in *Proc. IEEE ICCV,* Seoul, Korea,2019, pp. 7919–7929.

[16] A. Wu, W. S. Zheng and J. H. Lai, "Unsupervised person re-identification by camera-aware similarity consistency learning," in *Proc. IEEE ICCV,* Seoul, Korea, 2019, pp. 6922–6931.

[17] Y. Fu, Y. C. Wei, Y. Q. Zhou, H. H. Shi, G. Huang, X. C. Huang, Z. Q. Yao and T. Huang, "Horizontal pyramid matching for person re-identification," in *Proc. AAAI,* Hawaii, USA, 2019, pp. 8259–8302.

[18] F. Zheng, C. Deng, X. Sun, X. Jiang, X. Guo, Z. Yu, F. Huang, and R. Ji, "Pyramidal person re-identification via multi-loss dynamic training,"in *Proc. IEEE CVPR,* California, USA, 2019, pp. 8514–8522.

[19] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang and Q. Tian, "Scalable person re-identification: a benchmark," in *Proc. IEEE ICCV,* Santiago, Chile,2015, pp. 1116–1124.

[20] 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. IEEE ECCV,* Amsterdam, The Netherlands,2016, pp. 17–35.

[21] Z. D. Zheng, L. Zheng and Y. Yang, "Unlabeled samples generated by gan improve the person re-identification baseline in vitro," in *Proc. IEEE ICCV,* Venice, Italy, 2017, pp. 3754–3762.

[22] M. Ye, J. B. Shen, G. J. Lin, T. Xiang, L. Shao and S. C. H. Hoi, "Deep learning for person re-identification:a survey and outlook,"arXiv preprint arXiv:2001.01493,2020. [Oneline]. Available: https://arxiv.org/abs/2001.01493.

[23] Y. T. Lin, L. X. Xie, Y. Wu, C. G. Yan, Q. Tian, "Unsupervised Person Re-identififfication via Softened Similarity Learning," in *Proc. IEEE CVPR,* 2020, pp. 3390–3399.

[24] D. K. Wang, S. L. Zhang, "Unsupervised Person Re-identification via Multi-label Classification," in *Proc. IEEE CVPR,*, 2020, pp. 10981–10990.

[25] Y. T. Lin, X. Y. Dong, L. Zheng, Y. Yan and Y.Yang, "A bottom-up clustering approach to unsupervised person re-identififfication," in *Proc. AAAI,* 2019, pp. 8738–8745.

[26] K.W. Zeng, M. N. Ning, Y. H. Wang and Y. Guo, "Hierarchical clustering with hard-batch triplet loss for person re-identififfication," in *Proc. CVPR,* 2020, pp. 13657–13665.

[27] Z. L. Ji, X. L. Zou, X. H. Lin, X. liu, T. J. Huang and S. Wu, "An attention-driven two-stage clustering method for unsupervised person re-identififfication," in *Proc. ECCV,* 2020.

[28] Y. P. Zhai, S. J. Lu, Q. X. Ye, X. B. Shan, J. Chen, R. G. Ji and Y. H. Tian, "AD-Cluster: augmented discriminative clustering for domain adaptive person re-identification," in *Proc. CVPR,* 2020, pp. 9021–9030.

[29] F. X. Yang, L. Li, Q. Z. Zhong, Z. M. Luo, X. Sun, H. Chen, X. W. Guo, F. Y. Huang, R. R. Ji and S. Z. Li, "Asymmetric co-teaching for unsupervised cross-domain person re-identification," in *Proc. AAAI,* 2020, pp. 12597–12604.

[30] F. Zhao, S. C. Liao, G. S. Xie, J. Zhao, K. H. Zhang and L. Shao, "Unsupervised domain adaptation with noise resistible mutual-training for person re-identification," in *Proc. ECCV,* 2020

[31] H. Luo, W. Jiang, Y. Gu, F. Liu, X. Liao, S. Lai, and J.Gu,"A strong baseline and batch norm neuralization neck for deep person re-identification," arXiv preprint arXiv:1906.08332,2019. [Oneline]. Available: https://arxiv.org/abs/1906.08332.

[32] Y. Sun, Q. Xu, Y. Li, C. Zhang,Y. Li,S. Wang, and J.Sun, "Perceive where to focus: learning visibility-aware part-level features for partial person re-identification," in *Proc. IEEE CVPR,* California, USA, 2019, pp.393–402.

[33] L. Zhao, X. Li, J. Wang, and Y. Zhuang, "Deeply-learned part-aligned representations for person re-identification," in *Proc. IEEE ICCV,* Venice, Italy, 2017, pp. 3219–3228.

[34] C. Su, J. Li, S. Zhang, J. Xing, W. Gao, and Q. Tian, "Pose driven deep convolutional model for person re-identification," in *Proc. IEEE ICCV,* Venice, Italy, 2017, pp. 3960–3969.

[35] X. S. Chen, C. M. Fu, Y. Zhao, F. Zheng, J. K. Song, R. R. J and Y. Yhang "Salience-guided cascaded suppression network for person re-identification," in *Proc. IEEE CVPR,* Virtual ,2020, pp. 3300–1296.

[36] Y. F. Sun, L. Zheng, Y.Yang, Q. Tian and S. J. Wang "Beyond part models: person retrieval with refined part pooling," in *Proc. IEEE ECCV,* Munich, Germany, 2018, pp. 501–518.

[37] Q.Yang ,H. X. Yu, A. Wu,and W. S. Zheng,"Patch-based discriminative feature learning for unsupervised person re-identification,"in *Proc. IEEE CVPR,* California, USA, 2019, pp.3633–3642.

[38] Y. Fu, Y. C. Wei, G. S. Wang, Y. Q. Zhou, H. H. Shi and T. S. F. Huang, "Self-similarity grouping: a simple unsupervised cross domain adaptation approach for person re-identification," in *Proc. IEEE ICCV,* Seoul, Korea,2019, pp. 6112–6121.

[39] B. Sun and K. Saenko, "Deep coral: correlation alignment for deep domain adaptation," in *Proc. IEEE ECCV,* Amsterdam, The Netherlands, 2016, pp. 443–450.

[40] X. Zhang, F. X. Yu, S. F. Chang, S. Wang, "Deep transfer network: Unsupervised domain adaptation", arXiv preprint arXiv:1503.00591. [Oneline]. Available: https://arxiv.org/abs/1503.00591

[41] J. C. Wang, J. H. Lai, J. M. Wang, W. Q, liang and G. G. Wang, "Smoothing adversarial domain attack and p-memory reconsolidation for cross-domain person re-identification," in *Proc. IEEE CVPR,* Vitural,2020, pp. 10568–10577.

[42] H. H. Fan , Z. Zheng and Y. Yang, "Unsupervised person re-identification: clustering and fine-tuning," arXiv preprint arXiv:1705.10444,2017. [Oneline]. Available: https://arxiv.org/abs/1705.10444.

[43] J. M. Lv , W. H. Chen, Q. Li and C. Yang, "Un-supervised cross-dataset person re-identification by transfer learning of spatial-temporal patterns," in *Proc. IEEE CVPR,* Salt Lake City, USA, 2018, pp. 7948–7956.

[44] Y. Zhou, X. D. Yand, Z. D. Yu, B. Kumar and J. Kautz, "Joint disentangling and adaptation for cross-domain person re-identification," in *Proc. IEEE ECCV,* oneline, 2020.

[45] H. J. Wang, G. R. Wang, Y. Li, D. Y. Zhang and L. Lin, "Transferable, controllable, and inconspicuous adversarial attacks on person re-identification with deep mis-ranking," in *Proc. IEEE CVPR,* Vitural,2020, pp. 342–351.

[46] L. Qi, L. Wang, J. Huo, L. Zhou, Y. Shi, and Y. Gao,"A novel unsupervised camera-aware domain adaptation framework for person re-identification,"in *Proc. IEEE ICCV,* Seoul, Korea, 2019, pp.8080–8089.

[47] J. Song, Y. Yang, Y. Z. Song, T. Xiang, and T. M. Hospedales,"Generalizable person re-identification by domain-invariant mapping network,"in *Proc. IEEE CVPR,* California, USA, 2019, pp.719–728.

[48] S. Liao and L. Shao,"Interpretable and generalizable personre-identification with query-adaptive convolution and temporal lifting,"in *Proc. ECCV* , oneline,2020.
[49] Y. Ge, F. Zhu, D .Chen ,R. Zhao, and H. Li,"Self-paced contrastive learning with hybrid memory for domain adaptive object re-id,"arXiv preprint arXiv:2006.02713, 2018.[Oneline]. Available: https://arxiv.org/abs/2006.02713.
[50] Y. P. Zhai, Q. X. Ye, S. J. Lu, M. X. Jia, R. R. Ji and Y. H. Tian, "Multiple expert brainstorming for domain adaptive person re-identification," in *Proc. IEEE ECCV,* Vitural,2020.
[51] K. M. He, X. Y. Zhang, S. Q. Ren and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE CVPR,* Las Vegas, USA, 2016, pp. 770–778.
[52] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li and F. F. Li, "Imagenet: A large-scale hierarchical image database," in *Proc. IEEE CVPR,* Miami, FL, USA, 2009, pp. 248–255.
[53] A. Hermans, L. Beyer and B. Leibe,"In defense of the triplet loss for person re-identification," arXiv preprint arXiv:1703.07737, 2017. [Oneline]. Available: https://arxiv.org/abs/1703.07737.
[54] L. C. Song, C. Wang, L. F. Zhang, B. Du, Q. Zhang, C. Huang and X. G. Wang, "Unsupevised domain adaptive re-identification: Theory and practice," arXiv preprint arXiv:1807.11334, 2018.[Oneline]. Available: https://arxiv.org/abs/1807.11334.
[55] Z. Zhong, L. Zheng, D. L. Cao and S. Z. Li, "Reranking person re-identification with k-reciprocal encoding," in *Proc. IEEE CVPR,* Miami, FL, USA, 2017, pp. 3652–3661.
[56] M. Ester, H. P. Kriegel, J. Sander and X. W. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. KDD,* 1996, pp. 226–231.
[57] S. Lin, H. L. Li, C. T. Li and A. C. Kot, "Multi-task mid-level feature alignment network for unsupervised cross-dataset person re-identification," in *Proc. BMVC,* Newcastle, UK,2018.
[58] H. X. Yu, W. S. Zheng, A. C. Wu, X. W. Guo, S. G. Gong and J. H. Lai, "Invariance matters: Exemplar memory for domainadaptive person re-identification," in *Proc. IEEE CVPR,* California, USA, 2019, pp. 2148–2157.

WENHUI DONG received her MS degree in Communication and Information Systems and PhD degree in Pattern Recognition and Machine Intelligence from Shandong University in 2006 and 2015. Now she is a professor at college of physics and electronic information in Dezhou University. She focuses on computer vision,pattern recognition, and image processing.

PEISHU QU received his B.S. degree in Electronic Engineering from Qufu Normal University in 2003 and M.S. degree in Communication and Information Systems from Civil Aviation University of China in 2009. He is now a professor in Dezhou University. His research interests include computer vision, pattern recognition and image processing.

CHUNSHENG LIU received M.S. degree and Ph.D. degree in Pattern Recognition and Machine Intelligence from Shandong University, Jinan, China in 2012 and 2016, respectively. He was a postdoctor and visiting researcher in University of Washington from 2018 to 2019. He is currently an associate professor at School of Control Science and Engineering, Shandong University. His research interests include pattern recognition, machine learning, intelligent transportation system, and bionic intelligence.

YANKE TANG received the PhD in University of Chinese Academy of Sciences in 2008 and visted YALE University as a visiting scholar from 2016 to 2017. Now he is a professor of Dezhou University in China. His research interests are astroinformatics and image processing and pattern recognition of the Sun.

NING GAI is a professor of Dezhou University. She received the PhD degree from Beijing Normal University in 2009 and visisted YALE University as a visiting scholar from 2016 to 2017. Her research interests include asteroseismology, astroinformatics and image processing and pattern recognition of the Sun.