Dispersion based Clustering for Unsupervised Person Re-identification

Guodong Ding[12]
guodong.ding@njust.edu.cn
Salman Khan[3]
salman.khan@anu.edu.au
Qingze Yin[12]
qingzeyin@njust.edu.cn
Zhenmin Tang[12]
Tzm.cs@njust.edu.cn

[1] Nanjing University of Science and Technology, Nanjing, China
[2] Collaborative Innovation Center of Social Safety Science and Technology Nanjing, China
[3] Australian National University ACT, Australia

Abstract

The cumbersome acquisition of large-scale annotations for person re-identification task makes its deployment difficult in real-world scenarios. It is necessary to teach models to learn without explicit supervision. This paper proposes a simple but effective clustering approach for unsupervised person re-identification. We explore a basic concept in statistics, namely *dispersion*, to achieve a robust clustering criterion. Dispersion reflects the compactness of a cluster when assessed within and reveals the separation when measured at the inter-cluster level. Based on this insight, we propose a Dispersion based Clustering (DBC) approach which performs better at discovering the underlying data patterns. The approach can automatically prioritize standalone data points and prevents poor clustering. Our extensive experimental results demonstrate that the proposed methodology outperforms the state-of-the-art unsupervised methods on person re-identification.

1 Introduction

Person re-identification (re-ID) aims at establishing the identity correspondence across non-overlapping cameras. It has various important applications such as intelligent person search or analysis in multi-camera streams. Extensive work has been done to address this problem in a supervised fashion, which has led to impressive results in recent years [23, 24, 27, 28, 30, 39, 47]. However, acquiring manual identity labels in complex scenes is a demanding job, due to which unsupervised solutions have also been investigated in the literature. Traditional unsupervised solutions are based on hand-crafted features [6, 14, 16], saliency analysis [31, 40] and dictionary learning [11]. These initial attempts for unsupervised learning resulted in much inferior performance compared to supervised models.

Clustering, as an essential data analysis tool, has also been studied in unsupervised person re-ID. Fan *et al*. [5] proposed to pretrain a CNN model on an external re-ID dataset and then applied a k-means clustering on the target dataset to progressively select high reliability data points for model update. One challenge for [5] is the right choice of magic number k

(*i.e*. the number of identities) in k-means clustering. To fully avoid dependence on an auxiliary person re-ID dataset, Lin *et al*. [15] proposed a bottom-up clustering approach which alternatively trains a CNN model and performs clustering, without any extra data source.

The bottom-up clustering approach in [15] is essentially a hierarchical clustering algorithm. Two main categories of such algorithms are *agglomerative* and *divisive* clustering. The hierarchical structure inherently indicates that the criterion used for cluster merging or division is crucial, which normally is defined as a (dis)similarity measure. In [15], the minimum distance between images in two clusters is used as similarity for merging criterion. However, this criterion can be problematic as it only considers one pair of images from two clusters, discarding other useful cues. Such a naive criterion can lead to elongated clusters which results in poor performance due to incorrect merging of distinct identity clusters.

Here, we attempt to tackle this important problem by exploring data dispersion in the feature space. We consider a clustering to be a good one if it follows two fundamental properties *i.e*., intra-cluster compactness and inter-cluster well-separation. In statistics, dispersion is the extent to which a distribution is stretched or squeezed, thus denoting the clustering quality. Low intra-dispersion and high inter-dispersion is the sign of a valid cluster and vice versa. Thus, we propose to employ this simple and elegant criterion as the merging rule.

Our contributions are three-fold: **(I)** We propose a dispersion based clustering approach for unsupervised person re-ID. This criterion considers both within cluster compactness and between clusters separation. **(II)** The criterion has two major advantages *i.e*., automatic prioritization of isolated data points for merging and prevention of poor clustering. **(III)** The experimental results demonstrate that our approach outperforms the state-of-the-art methods on both image-based and video-based person re-ID datasets.

2 Related Work

Top-performing deep architectures are trained on massive amounts of labeled data. Most existing re-ID models are trained with human annotated ID labels in a supervised mode [13]. Therefore, their deployment in real-world applications is usually hindered by lack of large-scale annotated training sets. Some unsupervised methods with hand-crafted features have been proposed in recent years [6, 10, 11, 12, 16, 17, 21, 31, 32, 37, 41]. However, they achieve inadequate re-ID performance when compared to the supervised learning methods. Specifically, [6] exploited the property of symmetry in person images to deal with view variances. To handle the illumination changes and cluttered background, Ma *et al*. [19] proposed to combine the Gabor filters and the covariance descriptor. Fisher Vector is explored in [20] to encode higher order statistics of local features.

In the absence of labeled data for a certain task, domain adaptation often provides an attractive option given that labeled data of similar nature but from a different domain are available. The main practice of adaptation is to align the feature distribution between the source and target domain[7, 18, 25, 26, 29]. Recently, some cross-domain transfer learning methods [22, 46] have been studied in the field of person re-ID to deal with the misalignment between identities among different datasets. To better bridge this gap, [33] first train with attributes on source domain and learn a joint feature representation of both identity and attribute. A hetero-homogeneous learning approached is introduced in [45] to align domain distributions. Besides, some work uses generative adversarial networks (GAN) to generate augmented images to reduce the dataset differences [3, 46]. [3] explore image self-similarity and cross-domain dissimilarity for a target domain image translation. While [46] exploited

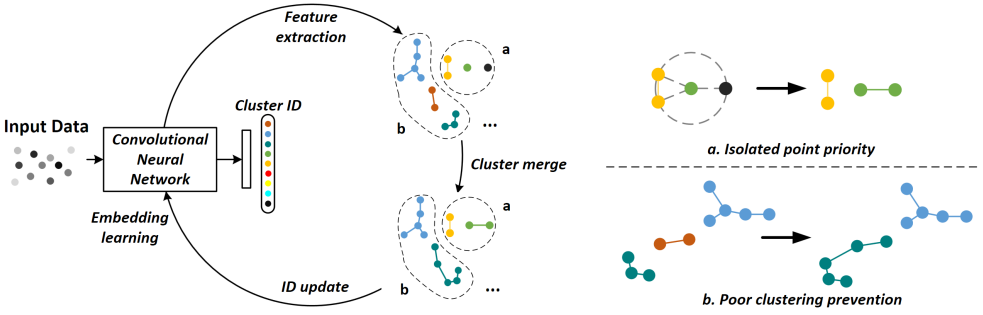


Figure 1: The overall framework for the unsupervised learning (*best viewed in color*). The left part shows the iterative process of our approach. The framework alternatively trains CNN model with cluster ID and performs cluster merging based on dispersion criterion. The right part exhibits two properties our dispersion criterion. (a) Isolated point priority. (b) Poor clustering prevention. (see Sec. 3.5)

camera-to-camera alignment to perform image translation. Those domain adaptation methods all focus on the label estimation of the target domain.

Unlike these methods, Fan *et al.* [8] combines domain transfer and clustering for unsupervised re-ID task. They first train the model on an external labeled dataset which is used as a good model initialization. After that, unlabeled data are progressively selected for training according to their credibility defined as the distance to cluster centroids. However, this work relies on a strong assumption about the total number of identities. Aside from these methods that requires auxiliary datasets or assumptions, [15] proposed to apply a bottom-up framework for clustering, which hierarchically combines cluster according to some criterion and achieved promising results. The merging in [15] is based on a very simple minimum distance criterion with a cluster size regularization term. Different from their work, our merging criterion exploits feature affinities within and between clusters, which also has a mutual promotion interaction with CNN model training process.

3 Dispersion based Clustering

We present a novel dispersion based clustering algorithm to perform un-supervised person re-ID. Our merging criterion helps in forming well-separated and compact clusters that in turn help in achieving better performance through our proposed self-learning strategy. We introduce our approach below (see Figure 1 for an overview).

3.1 Preliminaries

Given an unlabeled training set $\mathcal{D} = \{x_i\}_{i=1}^{N}$ containing N cropped person images, we aim to learn a feature embedding function $\phi(x_i;\theta)$ from $\mathcal{D}$ without any available annotations. The parameters θ are optimized iteratively using a proposed objective function. This feature extractor can later be applied to the gallery set $\{x_i^g\}_{i=1}^{N_g}$ and the query set $\{x_i^q\}_{i=1}^{N_q}$ to obtain their feature representations for a distance based retrieval. The distance between each pair of

images is defined as, $dist(x_i^q, x_i^g) = \|\phi(x_i^q;\theta) - \phi(x_i^g;\theta)\|$. For a higher distance based rank of a given pair, it is more likely that the pair belongs to the same identity.

Supervised learning provides person identity label y_i for each input image x_i. To learn the mapping between input and output, the feature embedding function is appended by a classifier $f(\phi;w)$ parameterized by w. Thus, $\phi(x_i;\theta)$ can be learnt by optimizing the following objective function:

$$\min_{\theta,w} \sum_{i=1}^{N} l(f(\phi(x_i^q;\theta);w), y_i) \tag{1}$$

where l is the cross-entropy (CE) loss for classification. One shortcoming of CE loss is it does not explicitly minimizes the intra-class distances. To this end, center loss is proposed that seeks to achieve within class compactness.

Similar to center loss [4, 34], repelled loss [15, 36] can act as a classifier f which has the ability to jointly consider inter-class and intra-class variances by computing probability based on the feature similarity as follows:

$$p(y|x,V) = \frac{\exp(V_y^T v/\tau)}{\Sigma_{j=1}^{N} \exp(V_j^T y/\tau)}, \tag{2}$$

where τ is a temperature parameter that controls the softness of probability distribution over classes, v is the l_2 normalized image feature obtained from $\phi(x;\theta)$, while V is a lookup table (LUT) containing the centroid feature of each class. This LUT is updated on the fly and can avoid exhaustive computation of feature extraction.

3.2 Learning Framework

The main challenge towards using the above framework for an unsupervised setting lies in automatic label assignment for unlabeled data. Here, clustering comes as a natural choice as it aims to group similar entities together in the same group. In this paper, we propose a novel dispersion based agglomerative clustering approach. The choice of affinity/dissimilarity measure between two clusters is the key to our proposed algorithm. In the task of person re-ID, which focuses on identifying images of the same identity, the inter and intra-cluster similarity should be considered for a reasonable merging. This requisite is fulfilled by a novel merging criterion used in our agglomerative clustering approach.

Given a cluster $\mathcal{C}$ scattered in feature space, we define its dispersion $d(\mathcal{C})$ as the average pairwise distance within cluster members:

$$d(\mathcal{C}) = \frac{1}{n} \sum_{i,j \in \mathcal{C}} dist(\mathcal{C}_i, \mathcal{C}_j), \tag{3}$$

where n is the cardinality of cluster $\mathcal{C}$. As such, the dispersion between any pair of clusters can be written as follows:

$$d(\mathcal{C}_a, \mathcal{C}_b) = \frac{1}{n_a n_b} \sum_{i \in \mathcal{C}_a, j \in \mathcal{C}_b} dist(\mathcal{C}_{a_i}, \mathcal{C}_{b_j}). \tag{4}$$

To jointly consider both intra- and inter-cluster dispersion, we have the dissimilarity between clusters $\mathcal{C}_a$ and $\mathcal{C}_b$ formulated as:

$$D_{ab} = d_{ab} + \lambda(d_a + d_b), \tag{5}$$

Algorithm 1 Dispersion based Clustering Approach

Input: Training data $\mathcal{D} = \{x_i\}_{i=1}^N$, merging percentage $m \in (0,1)$, trade-off parameter λ, CNN model $\phi(\cdot, \theta_0)$
Output: Optimized model $\phi(\cdot; \hat{\theta})$
Initialize label $\mathcal{Y} = \{y_i = i\}_{i=1}^N$, Cluster number $C = N$, merge batch $k = N * m$
while $C > k$ **do**
  Train model with $\{x_i\}$ and $\{y_i\}$ with Eq. (1)
  Calculate cluster dissimilarity matrix $\mathcal{P}(\mathcal{C})$
  for 1:k **do**
    Select candidate clusters according to Eq. (5) and merge them
    update matrix $\mathcal{P}(\mathcal{C})$ with Eq. (6) and Eq. (7)
    $C \leftarrow C - 1$
  end for
  Update $\mathcal{Y}$ with new cluster $\mathcal{C}$
  Evaluate Performance $Perf$ on validation set.
  if $Perf > Perf^*$ **then**
    $Perf^* = Perf$
    Best model = $\phi(\cdot; \hat{\theta})$
  end if
end while

where d_{ab} and d_a are used in place of $d(\mathcal{C}_a, \mathcal{C}_b)$ and $d(\mathcal{C}_a)$ for notation simplicity, and λ is the trade-off parameter between two components.

The former component d_{ab} in Eq. (5), dispersion between clusters, is a measure for cluster dissimilarity. Cluster with low dispersion should be considered for merging as features from the same identity should be close in feature space. The later component $d_a + d_b$, which is the sum of dispersion of both candidate clusters, serves as a regularizer to the former component. On one hand, it can help prioritize standalone data points for merging at the starting stages. On the other hand, this term can prevent escalating "poor" clustering as the high dispersion within-cluster can overbalance the inter-cluster term. In fact, this candidate cluster selection strategy controls the trade-off between the tendency to form spatially closer clusters ($\lambda \to 0$) or more compact clusters ($\lambda \to +\infty$).

3.3 Matrix Update

The input to this clustering process is the dissimilarity matrix $\mathcal{P}(\mathcal{C})$, also referred as the proximity matrix. It is an $C \times C$ matrix whose $(i,j)^{th}$ element equals the inter cluster dispersion $d(\mathcal{C}_i, \mathcal{C}_j)$ between $\mathcal{C}_i$ and $\mathcal{C}_j$. $\mathcal{P}(\mathcal{C})$ can be efficiently computed by first calculating an image pairwise distance matrix which is the outer product of stacked feature vectors obtained from deep networks.

At each clustering step, when two clusters are merged, the size of dissimilarity matrix $\mathcal{P}$ becomes $(C-1) \times (C-1)$. In one operation, two rows and columns of corresponding merged cluster $\mathcal{C}_a$ and $\mathcal{C}_b$ are deleted and a new row and a new column are added that contain the updated dissimilarity between the newly formed cluster $\mathcal{C}_q$ and an old cluster $\mathcal{C}_s$. The dissimilarity between $\mathcal{C}_q$ and $\mathcal{C}_s$ can be found using our dispersion definition, as follows:

$$d_{qs} = \frac{n_a}{n_a + n_b} d_{as} + \frac{n_b}{n_a + n_b} d_{bs}. \tag{6}$$

Correspondingly, the intra-cluster dispersion of C_q is written as:

$$d_q = \frac{n_a d_a + n_b d_b + n_a n_b d_{ab}}{n_a + n_b + n_a n_b}. \tag{7}$$

3.4 Network Update

The overall dispersion based clustering approach is presented in Figure 1. We begin our agglomerative clustering process with each data point x_i assigned a unique label y_i. With this initial label, we feed image-label pair (x_i, y_i) into the classification network and train it for a few epochs using error-backpropagation. Subsequently, clustering is performed, where top-k cluster pairs with least dissimilarity defined as Eq. (5) are considered to be merged. k is a pre-defined number of merging. When clusters are combined, images from both clusters will be assigned an identical label for next round of CNN training. The complete procedure of our unsupervised learning approach can be found in Algorithm 1.

3.5 Discussion

The regularization term. The combination of the second component in Eq. (5) brings two advantages to the clustering process: *1) Isolated point priority*. When performing clustering on a re-ID dataset, it is plausible to assume that there exists a balanced distribution of samples among clusters. Standalone points should have a higher priority at the beginning stage of clustering as they may be further pushed away from points of their own identity as the CNN is trained to separate them. The priority shifting happens when two merging options have identical inter dispersion d_{ab}, the standalone data point with less (zero) intra-cluster dispersion gets promoted. An illustration can be found in Figure 1(a). *2) Poor clustering prevention*. One disadvantage of the nesting property of agglomerative clustering is that there is no way to recover from a "poor" clustering that may have occured in previous levels of the hierarchy [8]. The addition of the regularization term helps to avoid this. Consider the case where a poor cluster formed in previous merging step, the high intra-cluster dispersion would prevent it from being selected for merging in following turns, albeit it may have high rankings in intra-cluster dispersion based merging list. An illustration can be found in Figure 1(b).

Comparison with close work. Our work shares a similar spirit as that of Bottom-up Clustering (BUC) [15] and adopts an agglomerative clustering framework for the task of unsupervised person re-ID. We differ substantially in terms of cluster merge criterion. Lin *et al.* [15] adopted minimum distance between cross cluster samples to measure their dissimilarity. It is known that the single linkage algorithm has a chaining effect, *i.e.*, the dissimilarity d_{qs} is obtained from d_{as} and d_{bs} whichever is smaller ($d_{qs} = \min\{d_{as}, d_{bs}\}$). This implies it has a tendency to favor elongated clusters. Stretched clusters may hinder next iteration of model training with repelled loss which favours compact groups. Based on the presumption that training samples are evenly distributed among identities, Lin *et al.* [15] proposed to use cluster cardinality as a diversity regularization term which can not fully address this problem. In contrast, our criterion works on the pairwise distance between individual data point which can better exploit the inter-cluster relations. Also, our criterion formulation can help in forming compact and well-separated clustering.

4 Experiments

4.1 Datasets

Market-1501 [42] consists of 1,501 identities and 32,688 labeled images, among which 12,936 images of 751 identities are used for training and 19,732 images of 750 identities are used for testing. **DukeMTMC-reID** [24] contains 36,411 labeled images of 1,404 identities. We use 702 identities for training and remaining for testing. Specifically, 16,522 training images, 2,228 query images and 17,661 gallery images are used. **MARS** [43] is a video-based dataset for person re-ID, which contains 17,503 video clips of 1,261 identities. The training set comprises of 625 identities while testing set has 636 identities. **DukeMTMC-VideoReID** [35] is derived from DukeMTMC dataset. It has 2,196 tracklets of 702 identities for training, 2,636 tracklets of 702 identities for testing.

4.2 Protocols

Training. To perform unsupervised learning on above mentioned person re-ID datasets, the training protocols are changed as follows. For image-based datasets, training split remains the same except for the removal of identity labels. Similarly, for video-based datasets, the training samples are the tracklets without identity labels. Note that no extra annotation information are used for model initialization or our unsupervised feature learning.

Evaluation. When the training is done, the CNN model is used as the feature extractor. The outputted activations from the penultimate layer of CNN model is used as the person descriptor, while the descriptor for a tracklet input is the average of its frame features. These person descriptors are then used for a Euclidean distance based retrieval. We evaluate our methods with ran-k and mean average precision (mAP). Rank-k accuracy represents the retrieval precision and the mAP value reflects the overall precision and recall rates.

4.3 Implementation details

In our experiments, we adopt ResNet-50 [9] as the backbone architecture with pre-trained weights on ImageNet[7]. A two layer fully connected layer is added on top of the penultimate layer of ResNet-50 for a smaller feature embedding learning. The last classification layer is implemented by Eq. (2), in which τ is set to be 0.1. All datasets share the exact same set of hyperparameters if not specified. For CNN model training, we set the total training epoch is to be 20, batch size to be 16, dropout rate to be 0.5, m to be 0.05. The CNN model is optimized by Stochastic Gradient Descent (SGD) with momentum set to 0.9. Learning rate for parameters is initialized to 0.1 and decreased to 0.01 after 15 epochs. For cluster merging, the trade-off parameter λ in Eq. (5) is set to be 0.1.

4.4 Algorithm Analysis

We show the significance of each component of our clustering algorithm in Table 1.

The effectiveness of the inter-cluster dispersion term. We evaluate the effectiveness of our inter-cluster dispersion term by comparing to a very close work BUC [15]. For fair comparison, we report results of BUC without its regularization term in the first row of Table 1, denoted by BUC$^-$. The results when only inter-cluster dispersion is used are shown in third row, denoted by DBC$^-$. Across all four datasets, DBC outperforms BUC by a large

Methods	Market-1501		DukeMTMC-reID		MARS		DukeMTMC-VideoReID	
	rank-1	mAP	rank-1	mAP	rank-1	mAP	rank-1	mAP
BUC$^-$ [15]	62.9	33.8	41.3	22.5	55.5	31.9	60.7	50.8
BUC [15]	66.2	38.3	47.4	27.5	61.1	38.0	69.2	61.9
DBC$^-$	66.2	38.7	48.2	27.5	59.8	37.2	71.8	63.2
DBC	**69.2**	**41.3**	**51.5**	**30.0**	**64.3**	**43.8**	**75.2**	**66.1**

Table 1: The effectiveness of our dispersion based criterion and comparison with minimum distance criterion. $^-$ denotes the removal of regularization term, i.e., the *cluster cardinality* in BUC and *intra dispersion* in Ours. **Red**: the best performance.

Methods	Labels	Market-1501				DukeMTMC-reID			
		rank-1	rank-5	rank-10	mAP	rank-1	rank-5	rank-10	mAP
BOW[40]	**None**	35.8	52.4	60.3	14.8	17.1	28.8	34.9	8.3
OIM[36]	**None**	38.0	58.0	66.3	14.0	24.5	38.8	46.0	11.3
UMDL[24]	Transfer	34.5	52.6	59.6	12.4	18.5	31.4	37.6	7.3
PUL[8]	Transfer	44.7	59.1	65.6	20.1	30.4	46.4	50.7	16.4
EUG[33]	OneEx	49.8	66.4	72.7	22.5	45.2	59.2	63.4	24.5
SPGAN[6]	Transfer	58.1	76.0	82.7	26.7	46.9	62.6	**68.5**	26.4
TJ-AIDL[33]	Transfer	58.2	-	-	26.5	44.3	-	-	23.0
BUC[15]	**None**	**66.2**	**79.6**	**84.5**	**38.3**	**47.4**	**62.6**	68.4	**27.5**
DBC	**None**	**69.2**	**83.0**	**87.8**	**41.3**	**51.5**	**64.6**	**70.1**	**30.0**

Table 2: Experimental results on Market-1501 and DukeMTMC-reID. The "Labels" column lists the supervision used by that method. "Transfer" means it uses an external dataset with annotations. "OneEx" denotes that one labeled image per identity is used. "None" denotes no extra information is used. **Red**: the best performance. **Blue**: the second best performance.

margin of $\sim$6% in rank-1 accuracy and 7% in mAP. This performance gap exists because the minimum distance criterion essentially forms stretched cluster. In contrast, ours uses average pairwise distance which considers wider context. Notably, our model without the second term can achieve comparable results to that of full BUC model.

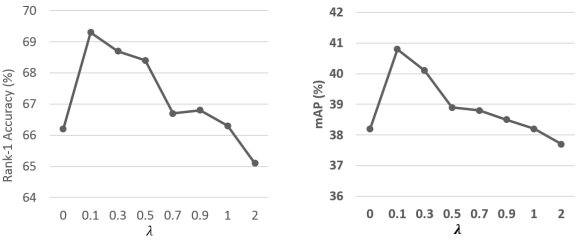
The effectiveness of the intra-cluster dispersion term. We further study the effect of the regularization term i.e., the intra-cluster dispersion. Our full model results (last row, Table 1) show that the regularization term helps to gain a performance boost. On Market-1501, the rank-1 accuracy is increased from 66.2% to 69.2% and mAP from 38.7% to 41.3%. A similar trend is observed across all datasets which advocates its effectiveness. With the two terms combined together, the model achieves the best performance.

Trade-off parameter. The regularization parameter λ in Eq. (5) balances importance of the intra-cluster and inter-cluster dispersion. We report results on Market-1501 dataset with varying λ values in Figure 2. It can be seen that the rank-1 accuracy first increases to its peak when $\lambda = 0.1$ and then experiences a decline as shown in Figure 3(a). A similar trend can be also found for mAP scores (Figure 3(b)). It is plausible since this parameter can be interpreted as the preference in candidate cluster selection which emphasizes more on selecting clusters that are spatially close in feature space when λ is relatively small, but more on selecting compact candidates as λ increases.

4.5 Comparison with state-of-the-art

We evaluate our approach on both image-based and video-based person re-ID datasets.

Figure 2: Parameter study on Market-1501 dataset. We set varying values for trade-off parameter λ and report Rank-1 accuracy and mAP score changes in (a) and (b), respectively. The best performance is achieved at $\lambda = 0.1$



(a) Rank-1 accuracy (b) mAP score

Methods	Labels	MARS				DukeMTMC-VideoReID			
		rank-1	rank-5	rank-10	mAP	rank-1	rank-5	rank-10	mAP
OIM[36]	**None**	33.7	48.1	54.8	13.5	51.1	70.5	76.2	43.8
DGM+IDE[37]	OneEx	36.8	54.0	-	16.8	42.3	57.9	69.3	33.6
Stepwise[17]	OneEx	41.2	55.5	-	19.6	56.2	70.3	79.2	46.7
RACE[33]	OneEx	43.2	57.1	62.1	24.5	-	-	-	-
DAL[1]	Camera	49.3	65.9	72.2	23.0	-	-	-	-
BUC[15]	**None**	61.1	**75.1**	**80.0**	38.0	69.2	81.1	**85.8**	61.9
EUG[35]	OneEx	**62.6**	74.9	-	**42.4**	**72.7**	**84.1**	-	**63.2**
DBC	**None**	**64.3**	**79.2**	**85.1**	**43.8**	**75.2**	**87.0**	**90.2**	**66.1**

Table 3: Results on MARS and DukeMTMC-VideoReID. The "Labels" column lists the supervision used by that method. "OneEx" denotes one labeled example per person is used. "Camera" denotes camera view information is provided. "None" denoted no extra information used. **Red**: the best performance. **Blue**: the second best performance.

Image-based Person Re-identification. Table 2 summarizes the state-of-the-art unsupervised person re-ID results on Market-1501 and DukeMTMC-reID datasets. On Market-1501, we achieve the best performance among all listed approaches with **rank-1 = 69.2%, mAP = 41.3%**. Among which, OIM [36] and BUC [15] are evaluated under the fully unsupervised setting. It can be seen that ours outperforms the state-of-the-art BUC by a margin of 3%. Similar performance improvements can be observed on DukeMTMC-reID dataset.

Performance of some domain adaption and one-shot learning approaches are also reported, *e.g.* TJ-AIDL [33] and EUG [35]. TJ-AIDL [33] trains with attribute labels to learn a robust embedding encoding extra attribute information which is transferable, while EUG [35] initializes model with one example labels and then progressively selects samples for training. In our experiment, we still surpass them by a relatively large margin (11% and 19.4% in rank-1 accuracy) even though external supervisions are used in their settings.

Video-based Person Re-identification. The comparisons with state-of-the-art algorithms on video-based person re-ID datasets, MARS and DukeMTMC-VideoReID are reported in Table 3. On DukeMTMC-VideoReID, we achieved **rank-1=75.2%, mAP=66.1%**, exceeding the counterpart BUC [15] by 6% and 4.2% in rank-1 accuracy and mAP, respectively. This demonstrates a more stable generalization ability of our proposed clustering algorithm to different data distributions. We also managed to outperform all other competitive methods on MARS dataset with **rank-1=64.3%, mAP=43.8%**. These results illustrate the effectiveness of our proposed approach.

5 Conclusion

In this paper, we proposed a dispersion based clustering approach for unsupervised person re-ID. On one hand, the proposed criterion considers both intra- and inter-cluster dispersion and can perform better clustering. The former dispersion term enforces compact clusters, while the latter ensures the separation between them. On the other hand, the criterion can handle isolated points and prevents poor clustering. The overall performance evaluations and ablation study illustrates the effectiveness of our proposed method.

References

[1] Yanbei Chen, Xiatian Zhu, and Shaogang Gong. Deep association learning for unsupervised video person re-identification. In *BMVC*, 2018.

[2] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Fei-Fei Li. Imagenet: A large-scale hierarchical image database. In *IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[3] Weijian Deng, Liang Zheng, Qixiang Ye, Guoliang Kang, Yi Yang, and Jianbin Jiao. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 994–1003, 2018.

[4] Guodong Ding, Shanshan Zhang, Salman Khan, Zhenmin Tang, Jian Zhang, and Fatih Porikli. Feature affinity based pseudo labeling for semi-supervised person re-identification. *IEEE transcations on Multimedia*, 2019.

[5] Hehe Fan, Liang Zheng, Chenggang Yan, and Yi Yang. Unsupervised person re-identification: Clustering and fine-tuning. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 14(4):83, 2018.

[6] Michela Farenzena, Loris Bazzani, Alessandro Perina, Vittorio Murino, and Marco Cristani. Person re-identification by symmetry-driven accumulation of local features. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2360–2367. IEEE, 2010.

[7] Yaroslav Ganin and Victor Lempitsky. Unsupervised domain adaptation by backpropagation. In *International Conference on Machine Learning*, pages 1180–1189, 2015.

[8] John C Gower. A comparison of some methods of cluster analysis. *Biometrics*, pages 623–637, 1967.

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[10] Furqan M Khan and Francois Bremond. Unsupervised data association for metric learning in the context of multi-shot person re-identification. In *IEEE International Conference on Advanced Video and Signal Based Surveillance*, pages 256–262, 2016.

[11] Elyor Kodirov, Tao Xiang, and Shaogang Gong. Dictionary learning with iterative laplacian regularisation for unsupervised person re-identification. In *Bristish Machine Vision Conference*, 2015.

[12] Elyor Kodirov, Tao Xiang, Zhenyong Fu, and Shaogang Gong. Person re-identification by unsupervised l_1 graph learning. In *European conference on computer vision*, pages 178–195. Springer, 2016.

[13] Minxian Li, Xiatian Zhu, and Shaogang Gong. Unsupervised person re-identification by deep learning tracklet association. In *European Conference on Computer Vision*, pages 737–753, 2018.

[14] Shengcai Liao, Yang Hu, Xiangyu Zhu, and Stan Z Li. Person re-identification by local maximal occurrence representation and metric learning. In *IEEE conference on computer vision and pattern recognition*, pages 2197–2206, 2015.

[15] Yutian Lin, Xuanyi Dong, Liang Zheng, Yan Yan, and Yi Yang. A bottom-up clustering approach to unsupervised person re-identification. In *AAAI Conference on Artificial Intelligence*, volume 2, 2019.

[16] Giuseppe Lisanti, Iacopo Masi, Andrew D Bagdanov, and Alberto Del Bimbo. Person re-identification by iterative re-weighted sparse ranking. *IEEE transactions on pattern analysis and machine intelligence*, 37(8):1629–1642, 2015.

[17] Zimo Liu, Dong Wang, and Huchuan Lu. Stepwise metric promotion for unsupervised video person re-identification. In *IEEE International Conference on Computer Vision*, pages 2429–2438, 2017.

[18] Mingsheng Long, Yue Cao, Jianmin Wang, and Michael I Jordan. Learning transferable features with deep adaptation networks. In *International Conference on International Conference on Machine Learning*, pages 97–105. JMLR.org, 2015.

[19] Bingpeng Ma, Yu Su, and Frédéric Jurie. Bicov: a novel image representation for person re-identification and face verification. In *British Machine Vision Conference*, pages 11–pages, 2012.

[20] Bingpeng Ma, Yu Su, and Frédéric Jurie. Local descriptors encoded by fisher vectors for person re-identification. In *European Conference on Computer Vision*, pages 413–422. Springer, 2012.

[21] Xiaolong Ma, Xiatian Zhu, Shaogang Gong, Xudong Xie, Jianming Hu, Kin-Man Lam, and Yisheng Zhong. Person re-identification by unsupervised video matching. *Pattern Recognition*, 65:197–210, 2017.

[22] Peixi Peng, Tao Xiang, Yaowei Wang, Massimiliano Pontil, Shaogang Gong, Tiejun Huang, and Yonghong Tian. Unsupervised cross-dataset transfer learning for person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[23] Chunfeng Song, Yan Huang, Wanli Ouyang, and Liang Wang. Mask-guided contrastive attention model for person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1179–1188, 2018.

[24] Yumin Suh, Jingdong Wang, Siyu Tang, Tao Mei, and Kyoung Mu Lee. Part-aligned bilinear representations for person re-identification. In *European Conference on Computer Vision (ECCV)*, pages 402–419, 2018.

[25] Baochen Sun and Kate Saenko. Deep coral: Correlation alignment for deep domain adaptation. In *European Conference on Computer Vision*, pages 443–450. Springer, 2016.

[26] Baochen Sun, Jiashi Feng, and Kate Saenko. Return of frustratingly easy domain adaptation. In *AAAI Conference on Artificial Intelligence*, 2016.

[27] Yifan Sun, Liang Zheng, Weijian Deng, and Shengjin Wang. Svdnet for pedestrian retrieval. In *IEEE International Conference on Computer Vision*, pages 3800–3808, 2017.

[28] Yifan Sun, Liang Zheng, Yi Yang, Qi Tian, and Shengjin Wang. Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline). In *European Conference on Computer Vision*, pages 480–496, 2018.

[29] Eric Tzeng, Judy Hoffman, Trevor Darrell, and Kate Saenko. Simultaneous deep transfer across domains and tasks. In *IEEE International Conference on Computer Vision*, pages 4068–4076, 2015.

[30] Rahul Rama Varior, Mrinal Haloi, and Gang Wang. Gated siamese convolutional neural network architecture for human re-identification. In *European conference on computer vision*, pages 791–808. Springer, 2016.

[31] Hanxiao Wang, Shaogang Gong, and Tao Xiang. Unsupervised learning of generative topic saliency for person re-identification. In *British Machine Vision Conference*, 2014.

[32] Hanxiao Wang, Xiatian Zhu, Tao Xiang, and Shaogang Gong. Towards unsupervised open-set person re-identification. In *IEEE International Conference on Image Processing*, pages 769–773. IEEE, 2016.

[33] Jingya Wang, Xiatian Zhu, Shaogang Gong, and Wei Li. Transferable joint attribute-identity deep learning for unsupervised person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2275–2284, 2018.

[34] Yandong Wen, Kaipeng Zhang, Zhifeng Li, and Yu Qiao. A discriminative feature learning approach for deep face recognition. In *European conference on computer vision*, pages 499–515. Springer, 2016.

[35] Yu Wu, Yutian Lin, Xuanyi Dong, Yan Yan, Wanli Ouyang, and Yi Yang. Exploit the unknown gradually: One-shot video-based person re-identification by stepwise learning. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 5177–5186, 2018.

[36] Tong Xiao, Shuang Li, Bochao Wang, Liang Lin, and Xiaogang Wang. Joint detection and identification feature learning for person search. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3415–3424, 2017.

[37] Mang Ye, Andy J Ma, Liang Zheng, Jiawei Li, and Pong C Yuen. Dynamic label graph matching for unsupervised video re-identification. In *IEEE International Conference on Computer Vision*, pages 5142–5150, 2017.

[38] Mang Ye, Xiangyuan Lan, and Pong C Yuen. Robust anchor embedding for unsupervised video person re-identification in the wild. In *European Conference on Computer Vision*, pages 170–186, 2018.

[39] Haiyu Zhao, Maoqing Tian, Shuyang Sun, Jing Shao, Junjie Yan, Shuai Yi, Xiaogang Wang, and Xiaoou Tang. Spindle net: Person re-identification with human body region guided feature decomposition and fusion. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1077–1085, 2017.

[40] Rui Zhao, Wanli Ouyang, and Xiaogang Wang. Unsupervised salience learning for person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3586–3593, 2013.

[41] Rui Zhao, Wanli Oyang, and Xiaogang Wang. Person re-identification by saliency learning. *IEEE transactions on pattern analysis and machine intelligence*, 39(2):356–370, 2017.

[42] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *IEEE International Conference on Computer Vision*, pages 1116–1124, 2015.

[43] Liang Zheng, Zhi Bie, Yifan Sun, Jingdong Wang, Chi Su, Shengjin Wang, and Qi Tian. Mars: A video benchmark for large-scale person re-identification. In *European Conference on Computer Vision*, pages 868–884. Springer, 2016.

[44] Zhedong Zheng, Liang Zheng, and Yi Yang. Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In *IEEE International Conference on Computer Vision*, pages 3754–3762, 2017.

[45] Zhun Zhong, Liang Zheng, Shaozi Li, and Yi Yang. Generalizing a person retrieval model hetero-and homogeneously. In *European Conference on Computer Vision*, pages 172–188, 2018.

[46] Zhun Zhong, Liang Zheng, Zhedong Zheng, Shaozi Li, and Yi Yang. Camstyle: a novel data augmentation method for person re-identification. *IEEE Transactions on Image Processing*, 28(3):1176–1190, 2019.

[47] Sanping Zhou, Jinjun Wang, Jiayun Wang, Yihong Gong, and Nanning Zheng. Point to set similarity based deep feature learning for person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3741–3750, 2017.