Unsupervised Group Re-identification via Adaptive Clustering-Driven Progressive Learning

Hongxu Chen[1], Quan Zhang[1], Jian-Huang Lai[1,2,3,4*], Xiaohua Xie[1,2,3,4]

[1]School of Computer Science and Engineering, Sun Yat-Sen University, Guangzhou 510006, China
[2]Pazhou Lab (HuangPu), Guangdong 510000, China
[3]Guangdong Key Laboratory of Information Security Technology, Guangzhou 510006, China
[4]Key Laboratory of Machine Intelligence and Advanced Computing, Ministry of Education, China
{chenhx87, zhangq48}@mail2.sysu.edu.cn, {stsljh, xiexiaoh6}@mail.sysu.edu.cn

Abstract

Group re-identification (G-ReID) aims to correctly associate groups with the same members captured by different cameras. However, supervised approaches for this task often suffer from the high cost of cross-camera sample labeling. Unsupervised methods based on clustering can avoid sample labeling, but the problem of member variations often makes clustering unstable, leading to incorrect pseudo-labels. To address these challenges, we propose an adaptive clustering-driven progressive learning approach (ACPL), which consists of a group adaptive clustering (GAC) module and a global dynamic prototype update (GDPU) module. Specifically, GAC designs the quasi-distance between groups, thus fully capitalizing on both individual-level and holistic information within groups. In the case of great uncertainty in intra-group members, GAC effectively minimizes the impact of non-discriminative features and reduces the noise in the model's pseudo-labels. Additionally, our GDPU devises a dynamic weight to update the prototypes and effectively mine the hard samples with complex member variations, which improves the model's robustness. Extensive experiments conducted on four popular G-ReID datasets demonstrate that our method not only achieves state-of-the-art performance on unsupervised G-ReID but also performs comparably to several fully supervised approaches.

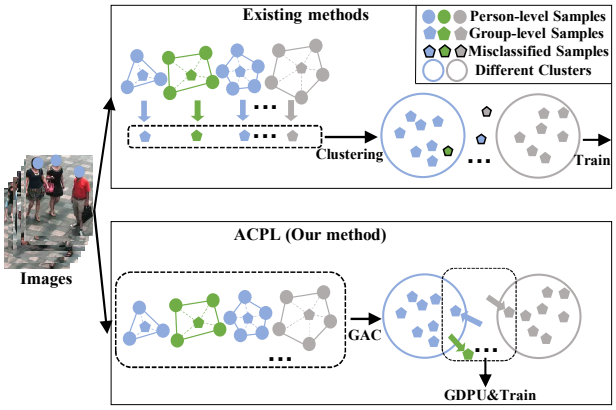


Figure 1: We compare existing methods with ACPL. Different colors represent the identities belonging to different groups. Existing clustering methods based on group features may create wrong labels with member changes, whereas our approach performs GAC to handle non-discriminative member features and get high-quality labels. Additionally, we explore hard samples during training with GDPU.

Introduction

Group re-identification (G-ReID) focuses on associating the group images containing the same members captured by different cameras. G-ReID usually deals with groups of 2 to 6 members, and considers group images with at least 60% of the same members as the same group class (Yan et al. 2020; Zhang et al. 2023). G-ReID plays an increasingly critical role in ensuring the safety of citizens by detecting and preventing heinous crimes such as child trafficking and kidnapping (Zhang et al. 2022a,b). In addition to the difficulty of expensive dataset labeling, the task of G-ReID is also challenged by the member variation, which means that the number of intra-group members may change due to members leaving or occlusion.

While supervised methods (Xu et al. 2019; Lin et al. 2019; Zhang et al. 2022a,b; Yan et al. 2020) dominate in existing approaches for the G-ReID task, they require a large amount of labeled data. Moreover, existing pure unsupervised methods are mainly based on hand-crafted features (Cai, Takala, and Pietikainen 2010; Zhu, Chu, and Yu 2016), and these methods exhibit subpar performance. Motivated by the popular method of unsupervised person re-identification (person ReID), which is driven by learning from pseudo-labels generated by clustering (Ester et al. 1996; Zhang et al. 2022c; Dai et al. 2022; Lan et al. 2023), we apply the same clustering methods to unsupervised G-ReID. However, the performances are not satisfactory. The main reasons for this limitation are as follows (the upper part of Fig.1): 1) Member variations can lead to distortions in the spatial distribution of group features, resulting in considerable instability when performing certain clustering. 2) During training, the utilization of a fixed momentum factor in memory bank updates may further limit the model's ability to adequately explore hard samples with complex member changes.

To address these problems, we propose our adaptive clustering-driven progressive learning (ACPL) method, as shown in the lower part of Fig.1. Specifically, we first pro-

*Corresponding Author.

pose a group adaptive clustering (GAC) module. GAC efficiently explores the information of group members, enabling better classification of hard samples with significant member changes during the clustering. Inspired by the Hausdorff Distance (Huttenlocher, Klanderman, and Rucklidge 1993), we design the quasi-distance vector between groups. Building on this, we are able to proficiently process non-discriminative member features across various groups, thereby optimizing and yielding high-quality pseudo-labels.

Furthermore, we employ a global dynamic prototype update (GDPU) strategy throughout the training iteration process. This strategy is motivated by the notion that prototypes associated with all pseudo-labels play a part in the update process for the current sample. To bolster the learning from hard samples, our GDPU strategy deploys a dynamic value to revise the prototypes associated with the current sample's pseudo-label. Adapting weights among complex member variations enables better learning of discriminative features from hard samples, significantly enhancing the model's robustness. Ultimately, experiments on four widely-used datasets demonstrate that our method achieves state-of-the-art (SOTA) performance on unsupervised G-ReID.

The contributions can be summarized as follows:

- We present adaptive clustering-driven progressive learning for unsupervised G-ReID. To solve the problems of member variations, group adaptive clustering (GAC) is proposed to handle non-discriminative member features in groups with significant uncertainty.
- To optimally utilize the hard samples derived from GAC, we propose the global dynamic prototype update (GDPU), which dynamically modifies the update weights during training and bolsters the robustness.
- Extensive experiments on four popular G-ReID datasets prove that our method is better than most previous unsupervised methods, and our method is competitive with state-of-the-art supervised methods.

Related Work

Group Re-Identification

The common methods in the early research of G-ReID are based on hand-crafted features: C-BRO (Zheng, Gong, and Xiang 2009), Covariance (Cai, Takala, and Pietikainen 2010), SBC (Salamon, Junior, and Musse 2015), BSC+CM (Zhu, Chu, and Yu 2016). Recently, deep learning-based methods have become mainstream, and most of them are supervised methods. Some methods (MGR (Lin et al. 2019), MACG (Yan et al. 2020)) mainly adopt the appearance features of groups. However, due to the increase or decrease of members and the changes of the relative positions in the group, the appearance features of groups may change dramatically, limiting the effectiveness of these methods. SOT (Zhang et al. 2022b) and 3DT (Zhang et al. 2022a) respectively propose new modeling methods that better solve the problems of group layout and membership changes. However, these fully supervised methods require fine-grained labeling of both the groups and members. To alleviate the burdensome process of manual labeling, methods under coarse-grained labeling (Xiao et al. 2018; Zhu et al. 2020; Huang et al. 2019a) are proposed and these methods do not require precise pedestrian labels. Meanwhile, other methods (Huang et al. 2019b, 2020; Yu et al. 2023) stem from a domain adaptation perspective, transferring the representation of individuals learned from an existing labeled ReID dataset to a target G-ReID domain. In contrast, our method requires neither identity labeling nor additional datasets, and is capable of handling member changes.

Unsupervised Person ReID

Methods for unsupervised person ReID can be divided into unsupervised domain adaptation (UDA) and pure unsupervised learning (USL). UDA methods (MMFA (Lin et al. 2018), TJAIDL (Wang et al. 2018), Invariance Matters (Zhong et al. 2019) and DMG-Net (Bai et al. 2021)) attempt to transfer knowledge from existing source data to unlabeled target data. For methods that do not require introducing a source domain dataset, state-of-the-art USL ReID pipelines adopt generating pseudo-labels and training deep neural networks. Some solutions are iterative clustering-based deep learning methods. Cluster Contrast (Dai et al. 2022) is proposed as a strong baseline, which stores feature vectors and computes contrast loss at the cluster level. Recent methods such as PPLR (Cho et al. 2022), ISE (Zhang et al. 2022c), and Purification (Lan et al. 2023) employ the Cluster Contrast in their training process. These methods rely on momentum updates (He et al. 2020) and often require a pre-set momentum factor, while we utilize information from the global prototypes for updating to learn from hard samples better.

Proposed Method

Our method alternates between a clustering step and a model updating step at each iteration. In this section, we first introduce our group adaptive clustering (GAC). Then we describe the initialization of the memory and update method in global dynamic prototype update (GDPU). Fig.2 illustrates our approach in detail.

Group Adaptive Clustering

Group adaptive clustering (GAC) aims to generate accurate pseudo-labels for groups, which handles the member variation on unsupervised G-ReID. In this process, a group is composed of pedestrians obtained through the cropping and processing of a single image. We treat each group as a set $G_i = \{p_1, p_2, \ldots, p_n\}$, where $p_1, p_2, \ldots, p_n$ are the extracted pedestrian features in one group. For the training set containing N groups, we put all the groups into a universal group set $\mathbf{U} = \{G_1, G_2, \ldots, G_i, \ldots, G_N\}$. Motivated by the Hausdorff Distance, we devise the one-way distance $d(p_m, G_i)$ from person to group in Eq.1:

$$d(p_m, G_i) = \min_{p \in G_i} \|p_m - p\|, \tag{1}$$

where p_m represents the pedestrian feature of the group other than G_i. $\|\cdot\|$ denotes the Euclidean norm of a vector.

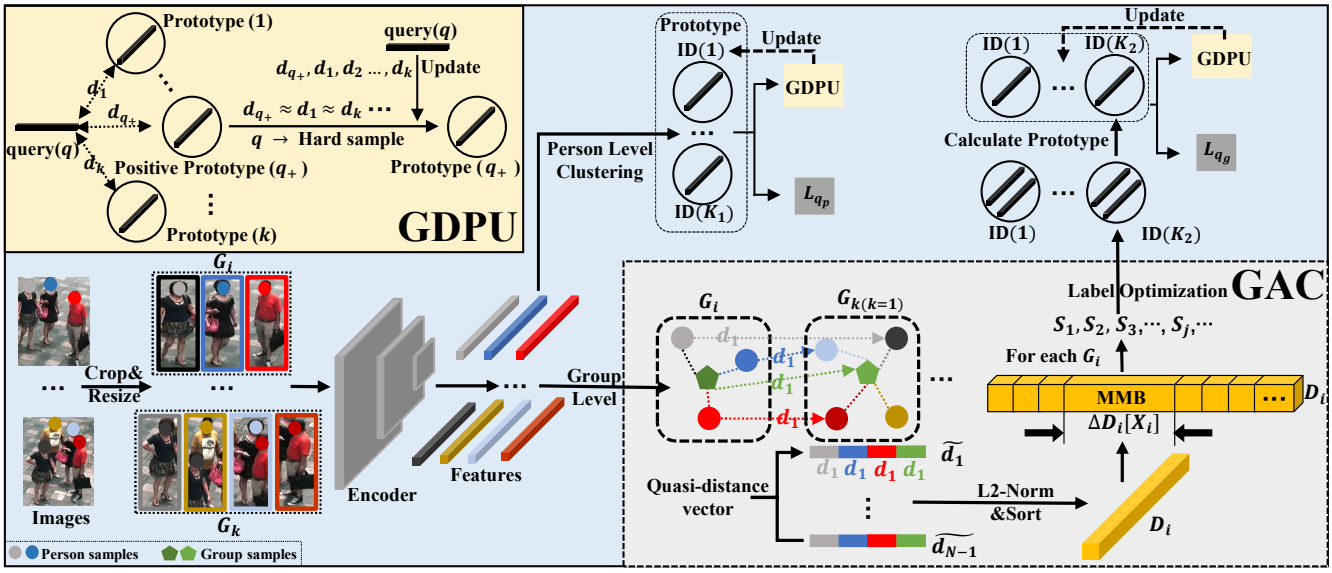


Figure 2: The overall architecture of our model. Our method starts with image input (bottom left) and consists of two stages: the first stage operates at the unsupervised person-level, while the second stage operates at the group-level. The circles or pentagons in different colors indicate samples with different features. d mainly represents the Euclidean distance between vectors.

For a group G_i with n members, we get the $(n+1)$ dimensional quasi-distance vector with respect to $G_k (k \neq i)$:

$$\tilde{d}(G_i, G_k) = (d(p_1, G_k), \cdots, d(p_n, G_k), \|G_{i_{ave}} - G_{k_{ave}}\|), \quad (2)$$

where $G_{i_{ave}}$ and $G_{k_{ave}}$ are the average pedestrian features of each member in group G_i and G_k, respectively. In a training set with N groups, each group G_i is associated with $N-1$ quasi-distance vectors. These vectors can be partitioned into two clusters, where the quasi-distance vectors closer to the origin are more likely to belong to groups having the same ID as G_i. In order to differentiate between different sets more efficiently, we first sort the quasi-distance vectors associated with G_i in ascending order according to their Euclidean norms. This enables to fully leverage both individual and holistic information within groups. To reduce computational overhead, we then select the smallest N_c norms to form a new vector D_i:

$$(\|\tilde{d}(G_i, G_{(1)})\|, \tilde{d}(G_i, G_{(2)})\|, \cdots, \|\tilde{d}(G_i, G_{(N_c)})\|), \quad (3)$$

the subscripts enclosed in parentheses indicate the result after reordering. To detect the mutation points, we perform a first-order discrete differential on D_i to obtain ΔD_i, and then locate the maximum margin boundary (MMB) X_i:

$$\Delta D_i[x] = D_i[x+1] - D_i[x], x = 1, 2, \cdots, N_c - 1, \quad (4)$$

$$X_i = \arg\max_x \{\Delta D_i[x]\}. \quad (5)$$

The outcome of Eq.5 may yield multiple X_i such that $\Delta D_i[X_i]$ takes the same value. In this case, we choose the smallest X_i. Additionally, $\Delta D_i[X_i]$ might be too minuscule to differentiate different clusters effectively. Therefore, we set a restriction on our MMB:

$$X_i = X_i \cdot \mathbb{I}\left\{\Delta D_i[X_i] > \frac{\epsilon_c \cdot (D_i[N_c] - D_i[1])}{N_c - 1}\right\}, \quad (6)$$

ϵ_c is a hyperparameter, and $\mathbb{I}\{\cdot\}$ is the indicator function. For each group $G_i (i = 1, 2, \cdots, N)$ with a positive MMB, we reserve the first X_i elements of D_i, identify the corresponding groups from the universal group set $\mathbf{U}$, and merge them with the original group G_i to form a new set S_i. We obtain the results of N preliminary sets $S_1, S_2, \cdots, S_i, \cdots, S_N (S_i = \emptyset \; if \; X_i = 0)$[1]. Next, we will proceed with the optimization of these preliminary sets and mitigate the impact of non-discriminative member features.

Suppose the minimal set consisting the non-empty preliminary sets is $\mathbf{S} = \{\tilde{S}_1, \tilde{S}_2, \cdots, \tilde{S}_j, \cdots, \tilde{S}_s\} (s \leq N)$. Let a set $\mathbf{Y} = \{y_1, y_2, \cdots, y_j, \cdots, y_s\}$ corresponding to the set $\mathbf{S}$, where the definition of y_j is: for a set $\mathcal{I}_j \subseteq \{1, 2, \cdots, N\}$ which satisfies:

$$\begin{cases} i \in \mathcal{I}_j, \text{if } \tilde{S}_j = S_i \\ i \notin \mathcal{I}_j, \text{if } \tilde{S}_j \neq S_i \end{cases}, \quad (7)$$

we get $y_j = |\mathcal{I}_j|$, and $|\cdot|$ is the cardinality of a set. Let a permutation of $y_1, y_2, \cdots, y_s$ be $y_{(1)} \geq y_{(2)} \geq \cdots \geq y_{(s)}$, we perform on the corresponding $\tilde{S}_{(1)}, \tilde{S}_{(2)}, \cdots, \tilde{S}_{(u)}, \cdots, \tilde{S}_{(s)}$ in turn:

- Update $\tilde{S}_{(1)}$ and get the first output set:

$$\tilde{S}_{(1)} = \begin{cases} \tilde{S}_{(1)}, \text{if } \left|\tilde{S}_{(1)}\right| \geq 2 \\ \emptyset, \text{otherwise} \end{cases}. \quad (8)$$

- Update $\tilde{S}_{(u)}, u \geq 2$ and get the u-th output set:

$$\tilde{S}_{(u)} = \begin{cases} \tilde{\mathbf{s}}_{(u)}, \text{if } |\tilde{\mathbf{s}}_{(u)}| \geq 2 \\ \emptyset, \text{otherwise} \end{cases}, \quad (9)$$

[1] $\emptyset$ denotes the empty set.

where $\tilde{\mathbf{s}}_{(u)} = \tilde{S}_{(u)} \setminus \bigcup_{k=1}^{u-1} \tilde{S}_{(k)}$.

For each group in a non-empty output set, we assign the same pseudo-label.

Global Dynamic Prototype Update

We apply global dynamic prototype update (GDPU) in our training after obtaining pseudo-labels at both the person-level and group-level. For clarity, we use subscripts p and g to represent the person-level and group-level training, respectively. For the training set, we put all the extracted pedestrian features into the all-person feature set $P = \{p_1, p_2, \ldots, p_M\}$. For the group-level initialization, we take the average of pedestrian features in each group and put the results of all groups into the all-group feature set $G = \{g_1, g_2, \ldots, g_N\}$. We then cluster P into K_1 clusters with DBSCAN (Ester et al. 1996) and G into K_2 clusters with our group adaptive clustering (GAC, the values for K_1 and K_2 are determined only after the clustering process is performed). For the person clusters and group clusters, we store each cluster's representation $\{\phi_p(1), \phi_p(2), \ldots, \phi_p(K_1)\}$ and $\{\phi_g(1), \phi_g(2), \ldots, \phi_g(K_2)\}$ in two memory-based feature dictionaries. We use the mean feature vectors of each cluster to initialize the cluster representation.

During training, we use a random identity sampler. At the person-level, a fixed number of pedestrian identities and the corresponding instances for each person identity are sampled from the training set in one minibatch, while the sampler at the group-level remains the same. For the feature q_p or q_g sampled in the query instance features set $\mathcal{Q}_p$ or $\mathcal{Q}_g$ in one iteration, we compute the loss at the cluster level by a cluster-wise contrastive loss:

$$L_{q_p} = -\log \frac{\exp\left(\phi_{p+}^T q_p/\tau\right)}{\sum_{k=0}^{K_1} \exp\left(\phi_p(k)^T q_p/\tau\right)}, \tag{10}$$

$$L_{q_g} = -\log \frac{\exp\left(\phi_{g+}^T q_g/\tau\right)}{\sum_{k=0}^{K_2} \exp\left(\phi_g(k)^T q_g/\tau\right)}, \tag{11}$$

where τ is the temperature. ϕ_{p+} and ϕ_{g+} are positive cluster representations corresponding to q_p and q_g, respectively.

To enable the model to more effectively capture discriminative features from samples with significant member variations, we utilize global prototype information to adjust the update weight. For the query features q_p and q_g, we use Eq.12 and Eq.13 to get the corresponding dynamic weights ξ_p and ξ_g:

$$\xi_p = \frac{\|q_p - \phi_p(k_p)\|}{\frac{\sum_{\lambda=1,\lambda\neq k_p}^{K_1} \|q_p - \phi_p(\lambda)\|}{K_1 - 1} + \|q_p - \phi_p(k_p)\|}, \tag{12}$$

$$\xi_g = \frac{\|q_g - \phi_g(k_g)\|}{\frac{\sum_{\lambda=1,\lambda\neq k_g}^{K_2} \|q_g - \phi_g(\lambda)\|}{K_2 - 1} + \|q_g - \phi_g(k_g)\|}, \tag{13}$$

where k_p and k_g are pseudo-labels for q_p and q_g, respectively. Then we use the ξ_p and ξ_g to update the person and group cluster representations. $\phi_p(k_p)$ is updated using Eq.14, while $\phi_g(k_g)$ is updated using Eq.15:

$$\frac{1}{|\mathcal{P}_{k_p}|} \left\{ \sum_{p_i \in \mathcal{P}_{k_p} \cap \mathcal{Q}_p} [\xi_p q_p + (1-\xi_p) p_i] + \sum_{p_i \in \mathcal{P}_{k_p} \setminus \mathcal{Q}_p} p_i \right\}, \tag{14}$$

$$\frac{1}{|\mathcal{G}_{k_g}|} \left\{ \sum_{g_i \in \mathcal{G}_{k_g} \cap \mathcal{Q}_g} [\xi_g q_g + (1-\xi_g) g_i] + \sum_{g_i \in \mathcal{G}_{k_g} \setminus \mathcal{Q}_g} g_i \right\}, \tag{15}$$

where the set $\mathcal{P}_{k_p}$ contains all person feature vectors in the k_p-th person-level cluster, and the set $\mathcal{G}_{k_g}$ contains all group feature vectors in the k_g-th group-level cluster. This adaptively updates weights of hard samples for better learning.

Experiment

Datasets and Implementation

We evaluate our proposed method on four G-ReID datasets: CSG (Yan et al. 2020), RoadGroup (Xiao et al. 2018), i-LIDS MCTS (Zheng, Gong, and Xiang 2009) and SYSU-Group (Mei et al. 2020). The CSG dataset contains 3,839 images, including 1,558 group classes. All images are collected from monitors and movies, and CSG adds an extra 5K group images as distractors in the gallery. The RoadGroup dataset contains 324 monitor images, including 162 group classes. We follow the division of these two datasets for training and testing as described in (Zhang et al. 2022b). The i-LIDS MCTS dataset contains 274 monitor images, including 64 group classes. The SYSU-Group dataset contains 7,071 group images with 208 group classes captured from 8 different cameras. We randomly and equally split the training and test sets of i-LIDS MCTS and SYSU-Group according to the protocol (Lin et al. 2019). These four datasets are commonly used real datasets in G-ReID tasks. We perform pedestrian detection on all datasets using PP-YOLOE (Xu et al. 2022) with a detection threshold of 0.7. If more than six pedestrians are detected in one image, we keep the six pedestrians with the highest resolution. We do not use any additional ReID datasets when training on each G-ReID dataset. We use Cumulative Matching Characteristics (CMC) at Rank-1 (R1), Rank-5 (R5), Rank-10 (R10), and mean Average Precision (mAP) as evaluation metrics.

Our model backbone is ResNet-50 (He et al. 2016), pretrained on ImageNet (He et al. 2016), and modified following Cluster-Contrast (Dai et al. 2022). We resize all cropped person images in the groups and enhance person samples in the training set as described in Cluster-Contrast. During training, we adopt Adam optimizer with weight decay 5e-4. The initial learning rate is set to 3.5e-4 and reduced to 0.1 of its previous value every 10 epochs. For person ReID, we train for 25 epochs with GDPU. At the beginning of each epoch, we first apply DBSCAN with the eps of 0.6 for pseudo-label assignment. During this phase, all pedestrians are sourced solely from the images in our training set, and no manual ID labels are used. We then use GAC to obtain group pseudo-labels and train for the G-ReID task for 25 epochs with the same settings as above. We set the values of ϵ_c and N_c to 3 and 10 respectively.

Methods	CSG				RoadGroup				i-LIDS MCTS			
	mAP	R1	R5	R10	mAP	R1	R5	R10	mAP	R1	R5	R10
MGR (Lin et al. 2019)	–	57.8	71.6	76.5	–	80.2	93.8	96.3	–	38.8	65.7	82.5
MACG (Yan et al. 2020)	–	63.2	75.4	79.7	–	84.5	95.0	96.9	–	45.1	70.4	84.9
DotSCN (Huang et al. 2020)	–	–	–	–	–	84.0	95.1	96.3	–	–	–	–
SVIGR (Mei et al. 2020)	–	–	–	–	89.2	87.8	92.7	–	42.1	46.2	71.8	–
BDFNet (Wang et al. 2022)	87.9	89.2	95.1	96.5	92.7	90.1	96.3	97.5	–	–	–	–
SOT (Zhang et al. 2022b)	90.7	91.7	96.5	97.6	91.3	86.4	96.3	98.8	–	–	–	–
3DT (Zhang et al. 2022a)	92.1	92.9	97.3	98.1	94.3	91.4	97.5	98.8	–	–	–	–
3DT+ (Zhang et al. 2022a)	94.4	95.1	97.7	98.6	94.8	93.8	97.5	98.8	–	–	–	–
C-BRO (Zheng et al. 2009)	–	10.4	25.8	37.5	–	17.8	34.6	48.1	–	23.3	54.0	69.8
Covariance (Cai et al. 2010)	–	16.5	34.1	47.9	–	38.0	61.0	73.1	–	26.5	52.5	66.0
BSC+CM (Zhu et al. 2016)	–	24.6	38.5	55.1	–	58.6	80.6	87.4	–	32.0	59.1	72.3
PREF (Lisanti et al. 2017)	–	19.2	36.4	51.8	–	43.0	68.7	77.9	–	30.6	55.3	67.0
LIMI* (Xiao et al. 2018)	–	–	–	–	–	72.3	90.6	94.1	–	37.9	64.5	79.4
DotGNN* (Huang et al. 2019a)	–	–	–	–	–	74.1	90.1	92.6	–	–	–	–
GCGNN* (Zhu et al. 2020)	–	–	–	–	–	81.7	94.3	96.5	–	41.9	68.1	86.9
PCPGU (Yu et al. 2023)	–	–	–	–	–	92.2	93.8	95.3	–	–	–	–
ACPL (Ours)	78.0	80.5	88.9	90.7	94.8	93.8	95.1	96.3	66.1	65.3	91.7	98.6

Table 1: Comparison with SOTA results on CSG, RoadGroup and i-LIDS MCTS datasets.

Methods	SYSU-Group			
	mAP	R1	R5	R10
GOG*(Matsukawa et al. 2016)	30.8	63.5	82.0	87.2
RPP*(Sun et al. 2018)	46.2	74.4	90.3	94.0
SVIGR*(Mei et al. 2020)	76.7	94.5	97.9	98.6
PCPGU (Yu et al. 2023)	–	97.4	98.6	99.1
ACPL (Ours)	85.7	96.5	98.4	99.1

Table 2: Comparison with SOTA results on SYSU-Group.

Perfomance

We evaluate our proposed method against the existing methods on four available G-ReID datasets to show the superiority of our method. As shown in Tab.1, the main body of this table is divided into two sections. The upper section includes fully supervised methods (MGR, MACG, DotSCN, SVIGR, BDFNet, SOT, 3DT, and 3DT+), which require both group and pedestrian identity labeling. The lower section includes supervised methods under coarse-grained labeling (LIMI, DotGNN, and GCGNN, these methods require group identity labeling, represented by *) and unsupervised methods (C-BRO, Covariance, BSC+CM, PREF, and PCPGU, these methods require no labels). In Tab.2, we compare with methods under coarse-grained labeling (GOG, RPP, SVIGR) and unsupervised method PCPGU. On these four datasets, compared to methods under coarse-grained labeling and unsupervised methods, our ACPL demonstrates superior or comparable performance. The earlier pure unsupervised methods based on hand-crafted features, such as C-BRO, Covariance, and PREF, are less effective. The use of global features cannot deal with problems such as member changes.

Some methods (DotGNN, DotSCN) have leveraged additional person ReID datasets to construct new groups for data augmentation, which to some extent have boosted performance, but these methods necessitate the introduction of extra data. Methods like PCPGU also requires pre-training on the MSMT17 dataset (Wei et al. 2018). Some supervised methods under coarse-grained labeling, generally strive to bolster performance at the group feature level directly. However, these techniques may overlook the negative influence of the non-discriminative member features within groups and the performance is unsatisfactory. The quasi-distance vector that we've designed is capable of exploring both holistic and individual features from a more comprehensive perspective. As the training progresses, different groups with the same identity are able to cluster together, even when there are complex internal member variations. Moreover, the accuracy of the clustering labels lays a solid foundation for us to effectively mine hard samples.

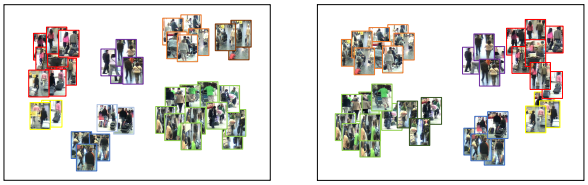
Compared with SOTA fully supervised methods, the performance of our method on RoadGroup is comparable on mAP/R1. Furthermore, our method outperforms existing methods on the i-LIDS MCTS. These results demonstrate that our method can effectively mine member associations and provide accurate pseudo-labels. In addition, our method has an advantage in efficient hard sample learning. However, there is a performance gap on the CSG dataset. This dataset contains a large number of images that include pedestrians who are not relevant to the current group. As a result, during unsupervised training, the model will inevitably learn the features of pedestrians outside of groups. This occurrence is less frequent on the other three datasets.

Effect of Group Clustering Method

To evaluate the effectiveness of our clustering approach, we utilize commonly used clustering methods, including Infomap (Rosvall and Bergstrom 2008), Spectral (Ng, Jordan, and Weiss 2002), K-means (Lloyd 1982), and DBSCAN (Ester et al. 1996) on the all-group feature set. To ensure fairness, we use the same settings in the first stage

	CSG		RoadGroup		i-LIDS	
	mAP	R1	mAP	R1	mAP	R1
InfoMap	69.1	71.6	85.5	82.7	55.2	54.7
Spectral	70.5	73.7	91.0	88.9	62.9	62.7
K-means	71.0	73.3	91.3	88.9	61.5	61.3
DBSCAN	68.4	71.1	87.3	85.2	53.9	53.3
DBSCAN+	71.6	74.2	89.1	86.4	60.2	58.7
GAC	**78.0**	**80.5**	**94.8**	**93.8**	**66.1**	**65.3**

Table 3: Experiments on clustering method on G-ReID.

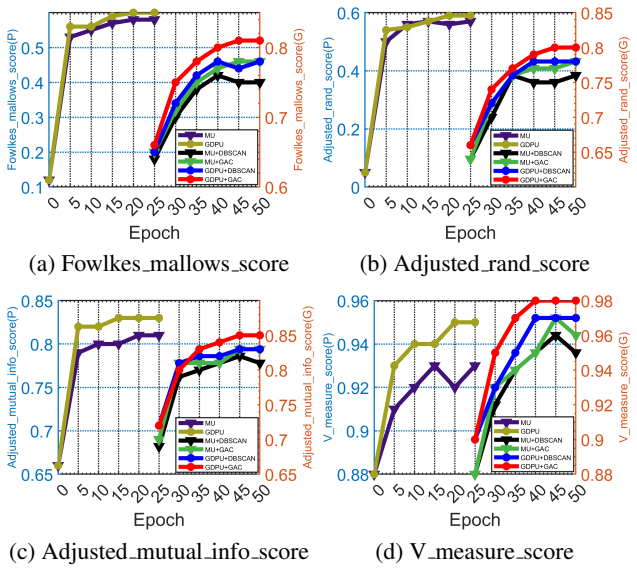


(a) DBSCAN (b) GAC

Figure 3: The t-SNE (Van and Hinton 2008) visualization of partial results embedded with DBSCAN and GAC on some samples. Different colors of image boxes represent different pseudo-labels (potential errors).

on person ReID. In the second stage, we set the eps value of DBSCAN to 0.6, and for Infomap, we chose two-level clustering. In Tab.3, our proposed GAC achieves the best performance among all the methods. Furthermore, we compare GAC with spectral and K-means clustering, which are sensitive to the number of clusters specified. When the number of clusters is set to the true total number of identities in the training set, spectral and K-means clustering show excellent performance in row 2 and row 3 of Tab.3. However, our method still outperforms these two methods, providing further evidence of its effectiveness. Additionally, we use the norm of Eq.2 as the distance metric between different samples in DBSCAN and take the average when asymmetry exists (DBSCAN+). Our quasi-distance vector effectively improves the clustering performance of DBSCAN, demonstrating its capability to handle potential member variations and obtain more precise pseudo-labels.

To further investigate the effectiveness of our method in improving pseudo-label quality for the G-ReID task, we conduct a visual analysis of partial groups on the i-LIDS MCTS dataset, utilizing DBSCAN or GAC during training. As shown in Fig.3, training strategy embedded with GAC effectively maps groups of the same ID with member variations closer in space (blue boxes) after multiple training epochs. GAC can adaptively adjust the MMB based on the number of members in the group, thereby improving the clustering effect for groups with great uncertainty that are challenging for DBSCAN. Additionally, GAC can make correct fine-grained distinctions for different groups that have members with high similarity (green boxes). For example, members from different groups may wear clothes in the same color or have similar spatial mapping positions. GAC is able to handle these groups well.



(a) Fowlkes_mallows_score (b) Adjusted_rand_score

(c) Adjusted_mutual_info_score (d) V_measure_score

Figure 4: Clustering quality over different epochs on CSG. In the first 25 epochs, we compare the clustering quality of GDPU and momentum update method in the baseline (MU) on person ReID task (P). In the last 25 epochs, in addition to comparing the effect of GDPU and MU on group clustering performance, we also compare the effect of DBSCAN and GAC on G-ReID (G).

We also assess the clustering quality at different training epochs on CSG, as shown in Fig.4. Four metrics are utilized, and higher scores indicate better performance. The comparison results of MU and GDPU will be discussed in a later section. It is evident that GAC has achieved superior performance on all metrics on G-ReID in the last 25 epochs under the same conditions of using MU or GDPU. The clustering quality of GAC steadily improves during the training process, while it may deteriorate in DBSCAN on G-ReID. The quasi-distance vector and the subsequent label optimization of GAC can refine the data distribution in the embedding space, reducing the noise in pseudo-labels from groups with high uncertainty.

Ablation Studies

Effect of Quasi-distance Vector We conduct experiments to compare the results of using the first n dimensions (individual features, IF) and only the $(n+1)$-th dimension (overall feature, OF) in Eq.2. We also compare using the feature distance of the most similar person between two groups as the distance metric directly (SF) (Mei et al. 2020). Tab.4 shows that the performance of using OF is the worst. OF cannot capture reliable fine-grained information, as OF considers only global features. On the other hand, if only individual features (IF or SF) in the group are used in the distance measure, the model will focus more on the features of person ReID. This may ignore the essential requirements of the original task G-ReID. On G-ReID, the association and member changes among group members are inseparable,

Method	CSG		RoadGroup		i-LIDS	
	mAP	R1	mAP	R1	mAP	R1
OF	55.0	58.5	80.2	76.5	53.8	53.3
SF	73.4	75.6	88.9	87.7	63.6	61.3
IF	76.3	78.2	92.4	90.1	65.5	64.0
OF+IF	**78.0**	**80.5**	**94.8**	**93.8**	**66.1**	**65.3**

Table 4: Ablation experiments on the quasi-distance vector.

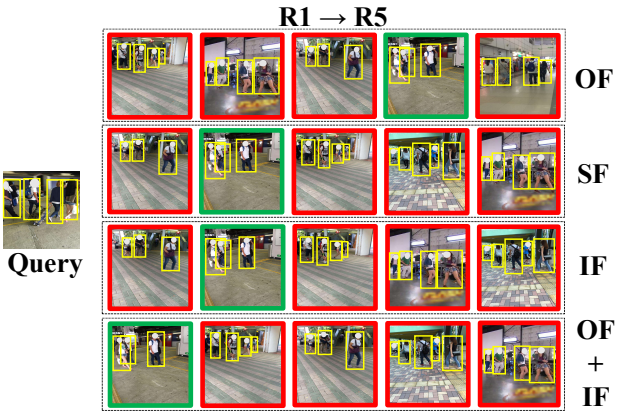
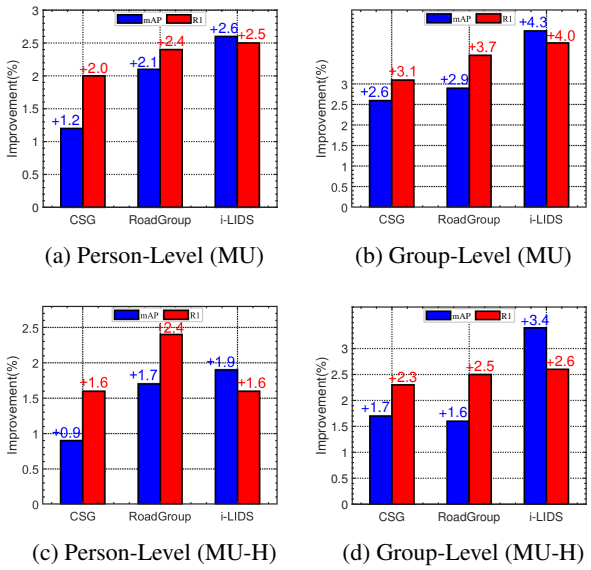


Figure 5: Visualization of top five retrieval results. Each row represents a different method used in training stage, where "OF+IF" is our method in GAC. Note that this query has only one correct image in the gallery. The green/red image box represents the correct/wrong matching.

and only by combining the individual and holistic features can we better handle the G-ReID task. This is also supported by the fact that GAC (IF+OF) achieves the best performance.

In Fig.5, we enumerate one visual retrieval example of different dimensions in the quasi-distance. When using OF, the retrieval tends to focus on images with the same number of pedestrians as the query, as shown in row 1 of Fig.5. This is because OF fails to address the problem of group member changes. On the other hand, if we only use IF as the distance measure, the similarity between two groups may depend only on the most similar pedestrian in each group (like SF), ignoring the overall group information. However, our method can retrieve these challenging samples correctly, as shown in row 4 of Fig.5.

Effect of Prototype Update Method When comparing with the momentum update method (MU), the momentum factor is set to the same value as Cluster-Contrast on person ReID and G-ReID experiments. For fairness, the clustering method for person ReID uniformly uses the DBSCAN in the Cluster-Contrast. Compared with the MU, Fig.4 shows that our GDPU achieves superior performance than MU on all metrics on person ReID during the first 25 epochs. In the last 25 epochs, we compare GDPU with MU under the same premise. All else being equal, our GDPU curves are basically above MU. Our method is effective in increasing the updated weight of hard samples that are farther away from



(a) Person-Level (MU) (b) Group-Level (MU)

(c) Person-Level (MU-H) (d) Group-Level (MU-H)

Figure 6: The improvement of GDPU compared with the best result on the momentum update method (MU) and MU-H. MU-H means the memory bank is updated using the hardest sample in a mini-batch.

the pseudo-label prototype. This allows the model to better learn discriminative features from complex groups with high member uncertainty.

The efficacy of our GDPU strategy is also revealed in Fig.6. Experimental results in Fig.6a and Fig.6b show that GDPU outperforms MU in terms of the final performance of the model, both at the person-level and at the group-level. This validates the effectiveness of our strategy for dynamic adjusting the update parameters of the group prototypes. In Fig.6c and Fig.6d, We also compare the method of utilizing the hardest sample in one mini-batch to update the memory bank (MU-H) (Zhang et al. 2022c). Compared with MU-H, GDPU enhances model robustness by adaptively handling groups with diverse members, leveraging its capability to comprehend a wider range of data distributions. Therefore, compared to MU-H, GDPU exhibits varying degrees of improvement on three datasets.

Conclusion

In this paper, we propose an adaptive clustering-driven progressive learning (ACPL) approach for unsupervised G-ReID. Our unsupervised method can effectively alleviate the expensive labeling problem of G-ReID datasets. Compared to the commonly used clustering methods, ACPL performs better in classifying groups with significant member variations. This lays the foundation for our subsequent network to better learn discriminative features from groups with great uncertainty. On unsupervised G-ReID, ACPL achieves satisfactory performance on the CSG, RoadGroup, i-LIDS MCTS, and SYSU-Group datasets.

Acknowledgments

This project was supported in part by the NSFC(62076258, U22A2095) and Guangdong Project(2020B1515120085).

References

Bai, Y.; Jiao, J.; Ce, W.; Liu, J.; Lou, Y.; Feng, X.; and Duan, L.-Y. 2021. Person30k: A dual-meta generalization network for person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2123–2132.

Cai, Y.; Takala, V.; and Pietikainen, M. 2010. Matching groups of people by covariance descriptor. In *IEEE International Conference on Pattern Recognition*, 2744–2747.

Cho, Y.; Kim, W. J.; Hong, S.; and Yoon, S.-E. 2022. Part-based pseudo label refinement for unsupervised person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7308–7318.

Dai, Z.; Wang, G.; Yuan, W.; Zhu, S.; and Tan, P. 2022. Cluster contrast for unsupervised person re-identification. In *Proceedings of the Asian Conference on Computer Vision*, 1142–1160.

Ester, M.; Kriegel, H.-P.; Sander, J.; Xu, X.; et al. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Knowledge Discovery and Data Mining*, volume 96, 226–231.

He, K.; Fan, H.; Wu, Y.; Xie, S.; and Girshick, R. 2020. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9729–9738.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 770–778.

Huang, Z.; Wang, Z.; Hu, W.; Lin, C.-W.; and Satoh, S. 2019a. DoT-GNN: Domain-transferred graph neural network for group re-identification. In *Proceedings of the 27th ACM International Conference on Multimedia*, 1888–1896.

Huang, Z.; Wang, Z.; Hung, T.-Y.; Satoh, S.; and Lin, C.-W. 2019b. Group re-identification via transferred representation and adaptive fusion. In *IEEE Fifth International Conference on Multimedia Big Data*, 128–132.

Huang, Z.; Wang, Z.; Tsai, C.-C.; Satoh, S.; and Lin, C.-W. 2020. DotSCN: Group re-identification via domain-transferred single and couple representation learning. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(7): 2739–2750.

Huttenlocher, D. P.; Klanderman, G. A.; and Rucklidge, W. J. 1993. Comparing images using the Hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(9): 850–863.

Lan, L.; Teng, X.; Zhang, J.; Zhang, X.; and Tao, D. 2023. Learning to purification for unsupervised person re-identification. *IEEE Transactions on Image Processing*.

Lin, S.; Li, H.; Li, C.-T.; and Kot, A. C. 2018. Multi-task mid-level feature alignment network for unsupervised cross-dataset person re-identification. *arXiv preprint arXiv:1807.01440*.

Lin, W.; Li, Y.; Xiao, H.; See, J.; Zou, J.; Xiong, H.; Wang, J.; and Mei, T. 2019. Group reidentification with multi-grained matching and integration. *IEEE Transactions on Cybernetics*, 51(3): 1478–1492.

Lisanti, G.; Martinel, N.; Del Bimbo, A.; and Luca Foresti, G. 2017. Group re-identification via unsupervised transfer of sparse features encoding. In *Proceedings of the IEEE International Conference on Computer Vision*, 2449–2458.

Lloyd, S. P. 1982. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2): 129–137.

Matsukawa, T.; Okabe, T.; Suzuki, E.; and Sato, Y. 2016. Hierarchical gaussian descriptor for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1363–1372.

Mei, L.; Lai, J.; Feng, Z.; and Xie, X. 2020. From pedestrian to group retrieval via siamese network and correlation. *Neurocomputing*, 412: 447–460.

Ng, A. Y.; Jordan, M. I.; and Weiss, Y. 2002. On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems*, 849–856.

Rosvall, M.; and Bergstrom, C. T. 2008. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, 105(4): 1118–1123.

Salamon, N. Z.; Junior, J. C. J.; and Musse, S. R. 2015. A user-based framework for group re-identification in still images. In *IEEE International Symposium on Multimedia*, 315–318.

Sun, Y.; Zheng, L.; Yang, Y.; Tian, Q.; and Wang, S. 2018. Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline). In *Proceedings of the European Conference on Computer Vision*, 480–496.

Van, d. M. L.; and Hinton, G. 2008. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(11).

Wang, J.; Zhu, X.; Gong, S.; and Li, W. 2018. Transferable joint attribute-identity deep learning for unsupervised person re-identification. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2275–2284.

Wang, Y.; Zhang, Q.; Lai, J.; Xie, X.; and Dong, J. 2022. Learning Bi-directional Feature Propagation with Latent Layout Modeling for Group Re-identification. In *2022 26th International Conference on Pattern Recognition*, 907–913. IEEE.

Wei, L.; Zhang, S.; Gao, W.; and Tian, Q. 2018. Person transfer gan to bridge domain gap for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 79–88.

Xiao, H.; Lin, W.; Sheng, B.; Lu, K.; Yan, J.; Wang, J.; Ding, E.; Zhang, Y.; and Xiong, H. 2018. Group re-identification: Leveraging and integrating multi-grain information. In *Proceedings of the 26th ACM international conference on Multimedia*, 192–200.

Xu, Q.; Yang, H.; Chen, L.; and Zhai, G. 2019. Group re-identification with hybrid attention model and residual distance. In *IEEE International Conference on Image Processing*, 1217–1221.

Xu, S.; Wang, X.; Lv, W.; Chang, Q.; Cui, C.; Deng, K.; Wang, G.; Dang, Q.; Wei, S.; Du, Y.; et al. 2022. PP-YOLOE: An evolved version of YOLO. *arXiv preprint arXiv:2203.16250*.

Yan, Y.; Qin, J.; Ni, B.; Chen, J.; Liu, L.; Zhu, F.; Zheng, W.; Yang, X.; and Shao, L. 2020. Learning multi-attention context graph for group-based re-identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(6): 7001–7018.

Yu, L.; Huang, S.; Lai, J.; and Feng, Z. 2023. Patch-based camera-aware person-to-group learning and group similarity strategy for unsupervised group re-identification. *Neurocomputing*, 552: 126565.

Zhang, Q.; Dang, K.; Lai, J.-H.; Feng, Z.; and Xie, X. 2022a. Modeling 3D layout for group re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7512–7520.

Zhang, Q.; Lai, J.; Xie, X.; and Chen, H. 2023. A summary on group re-identification. *Journal of Image and Graphics*, 28(5): 1225–1241.

Zhang, Q.; Lai, J.-H.; Feng, Z.; and Xie, X. 2022b. Uncertainty modeling with second-order transformer for group re-identification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 3318–3325.

Zhang, X.; Li, D.; Wang, Z.; Wang, J.; Ding, E.; Shi, J. Q.; Zhang, Z.; and Wang, J. 2022c. Implicit sample extension for unsupervised person re-Identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7369–7378.

Zheng, W.-S.; Gong, S.; and Xiang, T. 2009. Associating groups of people. In *British Machine Vision Conference*, volume 2, 1–11.

Zhong, Z.; Zheng, L.; Luo, Z.; Li, S.; and Yang, Y. 2019. Invariance matters: Exemplar memory for domain adaptive person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 598–607.

Zhu, F.; Chu, Q.; and Yu, N. 2016. Consistent matching based on boosted salience channels for group re-identification. In *IEEE International Conference on Image Processing*, 4279–4283.

Zhu, J.; Yang, H.; Lin, W.; Liu, N.; Wang, J.; and Zhang, W. 2020. Group re-identification with group context graph neural networks. *IEEE Transactions on Multimedia*, 23: 2614–2626.