ANNOTATION EFFICIENT PERSON RE-IDENTIFICATION WITH DIVERSE CLUSTER-BASED PAIR SELECTION

Lantian Xue, Yixiong Zou, Peixi Peng, Yonghong Tian, Tiejun Huang

{ltxue, zoilsen, pxpeng, yhtian, tjhuang}@pku.edu.cn

ABSTRACT

Person Re-identification (Re-ID) has attracted great attention due to its promising real-world applications. However, in practice, it is always costly to annotate the training data to train a Re-ID model, and it still remains challenging to reduce the annotation cost while maintaining the performance for the Re-ID task. To solve this problem, we propose the Annotation Efficient Person Re-Identification method to select image pairs from an alternative pair set according to the fallibility and diversity of pairs, and train the Re-ID model based on the annotation. Specifically, we design an annotation and training framework to firstly reduce the size of the alternative pair set by clustering all images considering the locality of features, secondly select images pairs from intra-/inter-cluster samples for human to annotate, thirdly re-assign clusters according to the annotation, and finally train the model with the re-assigned clusters. During the pair selection, we seek for valuable pairs according to pairs' fallibility and diversity, which includes an intra-cluster criterion to construct image pairs with the most chaotic samples and the representative samples within clusters, an inter-cluster criterion to construct image pairs between clusters based on the second-order Wasserstein distance, and a diversity criterion for cluster-based pair selection. Combining all criteria above, a greedy strategy is develop to solve the pair selection problem and select pairs. Finally, the above clustering-selecting-annotating-reassigning-training procedure will be repeated until the annotation budget is reached. Extensive experiments on three widely adopted Re-ID datasets show that we can greatly reduce the annotation cost while achieving better performance compared with state-of-the-art works.

Index Terms— Annotation Efficient, Semi-Supervised Learning, Person Re-Identification

1. INTRODUCTION

Person Re-Identification (Re-ID) [1, 2, 3] has attracted great attention in the field of computer vision, which retrieves images of a person from non-overlapping cameras. However, the training of a Re-ID model always relies on a large number of annotated samples with high annotation cost [4].

To solve this problem, some works have been proposed.

For example, works [1, 2, 3] based on semi-supervised learning only require part of the dataset to be annotated, but the annotation cost is still high and the performance is still limited. Works [5, 4] inspired by active learning select few image pairs for human to annotate, but the annotation cost may still be related to the number of images. Therefore, it still remains challenging to reduce the annotating cost while maintaining the performance for Re-ID.

To achieve this goal, in this paper we utilize the verification based annotation, which verifies whether two images of each pair are from the same identity [5, 4]. Intuitively, valuable image pairs should be selected for annotation. Inspired by [6], the value of a image pair can be defined as its informativeness, which could be divided into fallibility and diversity.

Therefore, our aim is to select pairs from **the alternative pair set** according to the **fallibility** and **diversity** with an affordable budget, and train a Re-ID model based on the annotation. Observing this aim, we can find there are three important terms to design:

(1) **The alternative pair set**, which means the set of all possible image pairs, influences the upper bound of the number of pairs to annotate. In view that images closer in the embedding space locally will tend to have the same semantic identity, we design to reduce the size of the alternative pair set by first clustering all images, then representing each cluster with a single representative image, and finally constructing pairs between representative images and other images. With the clustering operation, the size of the alternative pair set is related to the number of clusters instead of images.

(2) **Fallibility**, which means samples misclassified by the model (feature extractor), could be more informative for each selected pair. Considering each cluster of images can be coarsely understood as an identity that the model classifies, the error that the model makes mainly lies in two aspects:

• Within each cluster, images not belonging to the same identity are grouped into the same cluster, which we term as the chaotic samples. To select pairs from this situation, we propose to construct image pairs between the representative sample and the most chaotic sample of each cluster, which we term as the **intra-cluster criterion**.

• Between clusters, clusters referring to the same identity are separated. To select pairs from this situation, we propose to construct image pairs between representative samples of

all clusters based on the second order Wasserstein distance between clusters, termed as the **inter-cluster criterion**.

(3) **Diversity**. As samples from the same cluster tend to have similar appearances, it measures the information provided by all selected pairs. Intuitively, if all selected pairs are related to a few portion of clusters, the redundancy of the annotated pairs would be high. We select pairs according to KL-Divergence, aiming to average the frequency of each cluster for annotation, which we term as the **diversity criterion**.

Combining all criteria above, a greedy strategy is developed to find image pairs from the reduced alternative pair set. After the annotation, for the intra-cluster situation, if images of a pair are verified to be from two identities, the cluster will be re-assigned to be separated into two clusters. In the meanwhile, for the inter-cluster situation, if images of a pair are annotated to belong to the same identity, the corresponding clusters will be re-assigned to be merged into one cluster. Finally, each (re-assigned) cluster will be assigned with a pseudo-class label, which will be utilized to train the model. The above clustering-selecting-annotating-reassigning-training procedure will be repeated until the annotation budget is reached.

To sum up, our contribution can be summarized as:

• An annotation and training framework is proposed to annotate pairs from intra-/inter-cluster samples and train the model with re-assigned clusters, which greatly reduces the annotation cost while achieving better performance compared with current state-of-the-art annotation efficient methods.

• A pair selection strategy is proposed to select the most valuable pairs according to pair fallibility and diversity, and a greedy strategy is developed to solve the selection problem.

• Extensive experiments on three widely adopted Re-ID datasets validated the effectiveness of the proposed method.

2. RELATED WORK

Person Re-Id with less labeled data:

Recently, some works try to use less labeled data to train the Re-ID model. For example, [7] learns a distance metric according to annotators' feed-backs. DRAL [4] proposed to balance the data annotating with model training by a human-in-the-loop active learning manners. ARR [8] proposed to select samples with high uncertainty. Compare with them, our diversity criterion is applied in cluster level with pairwise annotations, instead of the sample level measure. To better deal with the redundancy, MASS [5] introduced clustering procedure with assigned pseudo labels and a scatter step to mine and annotate wrongly clustered samples. However, the pair selections and annotations remains within the cluster so there may still be some redundancy.

Unsupervised Re-ID: Some works [9, 10, 11] apply its own individual as identity for each sample. Apart from them, the cluster pseudo-labeling strategy [12, 13, 14] is proven to be effective. C-contrast [15] also clusters and assigns pseudo labels in the unsupervised setting but it directly updates cluster

memory by single query when training. [16] proposed an unsupervised clustering method focusing on generating clusters based on the energy distance and cluster deviations. Some works such as [9] also merge clusters in Re-ID. Although with similar terms, our method differs in that we are searching for the most valuable pairs for annotating, which jointly optimizes for the fallibility and diversity.

3. METHOD

The propose the annotation efficient method includes procedures of clustering, pair-selection, annotation, cluster re-assignment, and training, as shown in Fig. 1.

3.1. Preliminaries

Given a person Re-ID dataset D, ordinarily it is divided into a training set D_{train} and a test set D_{test} with non-overlapping classes [17, 18], where each class refers to a person identity, and totally there are M identities of N person images in D_{train}. For testing, D_{test} is composed of queries and galleries for image-retrieval-like evaluation [17, 18].

In this work, we follow [4, 5] to start with the unsupervised Re-ID setting to assume we only have access to the unlabeled training samples, denoted as D_{train}^{un}, then aim to select data for human to annotate. The annotation unit, as is adopted in current works [4], is the verification of a pair of images (e.g., x_i and x_k) , which is marked as the verification function $v(x_i, x_k) \rightarrow \{0, 1\}$, where 1 denotes they are from the same identity. A deep network $F(x; \theta)$ with input x and parameters θ will be used for embedding extraction.

The pairs selected for annotation compose the selected pair set $\mathbb{P}_{sel} = \{..., (x_i, x_j), ...\}$ with a pre-defined maximum budget T (verification times). By repeating the selecting-annotating-training procedure for few iterations, the final trained model will be obtained, which will be evaluated on the test set. Following [4, 5], the goal of our proposed method is to maximize the performance of the model $F(x; \theta)$ while reducing the annotation cost as much as possible.

3.2. Optimization Objective

Intuitively, the most valuable pairs need to be selected for annotation. According to active learning [6, 19], the value of an image pair is defined as its informativeness, which can be divided into fallibility and diversity. The former item helps the model train more effectively, while the latter makes the model more robust. Thus we formulate the pair-selection task as an optimization problem, written as

$$\max O(\mathbb{P}_{sel}) = Fallibility + \alpha Diversity$$
$$s.t. |\mathbb{P}_{sel}| = T, \mathbb{P}_{sel} \in \mathbb{P}_{total} = {D_{train}^{un}}^2 \quad (1)$$

where α is a hyper-parameter, and T is the budget. This means we select a set of pairs $\mathbb{P}_{sel}$ that provide the most fallibility and diversity overall from an alternative pair set $\mathbb{P}_{total}$.

Next, we will discuss how to simplify, specify, and solve it in our framework for the person Re-ID task.
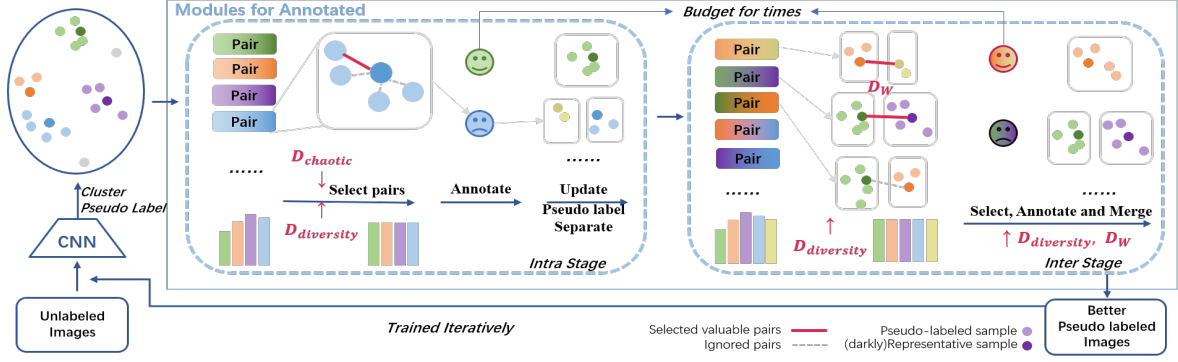
Fig. 1. Framework. We first cluster unlabeled images to reduce the size of the alternative pair set, then image pairs are selected according to the intra-cluster criterion, the inter-cluster criterion and the diversity criterion. After the annotation of selected pairs, clusters are re-assigned accordingly. Finally, the model will be trained with pseudo-class labels assigned for each cluster. The clustering-selecting-annotating-reassigning-training procedure will be repeated until the annotation budget is reached.

3.3. Annotation and Training Framework

The alternative pair set $\mathbb{P}_{total}$ means the set of all possible image pairs. In the worst case, the size of $\mathbb{P}_{total}$ could be as large as N × M if we naively annotating all possible pairs. However, great redundancy exists in this naive strategy. Considering that samples closer in the embedding space locally will tend to have the same semantic identity, we may not need to verify every pair. Therefore, same as [5], we first reduce the size of the alternative pair set by clustering all images.

In view that the clustering and pseudo-labeling based un-supervised learning methods [15] shows good performance in the Re-ID task, we consider to utilize this training procedure to take advantage of its simplicity and effectiveness.

Specifically, given an unlabeled training set, we first cluster all samples through each feature $f_i = F(x_i; \theta)$ to form the cluster set $\mathcal{C} = \{c^1, c^2, ..., c^{|C|}\}$, where the jth cluster is denoted as c^j. Later, the pair set $\mathbb{P}_{sel}$ will be selected based on these clusters, and clusters will be re-assigned according to the annotation on $\mathbb{P}_{sel}$, which will be illustrated in section 3.4 and 3.5. After the pair-selection and cluster-reassignment, all samples $x_i^j \in c^j$ will be re-assigned with a pseudo-label as $\widetilde{y}_i = j$, which will then be utilized to train $F(x; \theta)$. As the non-parametric classification loss has gained attention in person Re-ID recently, it inspires us to use the ClusterNCE [15] structure with its cluster based momentum memory update strategy. The loss function is written as

$$L_i = -log \frac{exp((f_i \cdot f^{\widetilde{y}_i})/\tau)}{\sum_{j=0}^{|C|} exp((f_i \cdot f^j)/\tau)} \qquad (2)$$

where f^j for cluster j is initialized by averaging $f_i(x_i \in c^j)$ and updated by momentum during each iteration [15].

3.4. Informative Pair Selection

After reducing the size of the alternative pair set $\mathbb{P}_{total}$, here we focus on the optimization objective in Eq. 1. We first introduce the fallibility term of the objective in this section, and then introduce the diversity term in section 3.5.

The CNN model $F(\cdot; \theta)$ is able to extract discriminative features after training. On that basis, we expect the annotation results of $\mathbb{P}_{sel}$ can further make up for deficiencies and correct mistakes. Intuitively, the more mistakes that $\mathbb{P}_{sel}$ can correct, approximately the more information it can provide for model training, and the more valuable they are. Therefore, two kinds of mistakes are set for pair selection in different stages:

- **intra-cluster stage**: correct mistakes that samples with different identities are clustered together.

- **inter-cluster stage**: correct mistakes that samples from the same identity are clustered into two groups. Considering the annotation budget T, with the clustering operation, we first pick up a sample r^j to represent the cluster c^j. Specifically, we select the sample closest to the cluster center f_{mean}^j to be the representative sample, which is

$$r^j = \arg\min_{x_i^j} \|f_i^j - f_{mean}^j\|_2, x_i^j \in c^j \qquad (3)$$

For each stage, we construct a subset $\mathbb{P}_{stage} \in \mathbb{P}_{total}$ to be the alternative pair set for the model to select pairs, written as

$$\mathbb{P}_{stage} = \begin{cases} \{(r^a, r^b) \mid c^a, c^b \in C\}, & Intra-cluster \\ \{(r^a, s^a) \mid c^a \in C\}, & Inter-cluster \end{cases} \qquad (4)$$

where s^a of c^a will be described later.

Then, in the proposed annotation and training framework, for a pair in $\mathbb{P}_{stage}$, we use the possibility that it may cause cluster re-assignment to approximate its information.

3.4.1. Intra-cluster stage: the chaotic degree distance

In the intra-cluster stage, the fallibility comes from whether a given sample belongs to the given cluster. For each sample $x_i \in c^j$, a typical way to measure whether it is from another cluster (i.e., the chaotic degree) is to measure the distance between f_i and f_{mean}^j. As the vertical position information is important in Re-ID, we also take this into consideration. In particular, before the final pooling layer of the CNN, we

record each channel about which vertical portion the maximum value comes from, denoted as g_i. Thus, the chaotic degree of x_i is defined as

$$d_{intra}(x_i^j) = \|f_i^j - f_{mean}^j\|_2 + \gamma\|g_i^j - g_{mean}^j\|_2, x_i^j \in c^j \quad (5)$$

where γ is a parameter for adjusting the proportions. This would be the intra-cluster criterion for the pair selection, which forms the first of fallibility in Eq. 1 for the intra-cluster stage. The chaotic distance of sample s^j is then be recorded as the chaotic degree of this pair.

3.4.2. Inter-cluster stage: the Wasserstein distance

In the inter-cluster stage, the fallibility comes from whether two clusters actually represent the same identity. Intuitively, the closer two clusters are, the more likely that they might be from the same identity. Therefore, we would need the measuring of the distance between clusters to express the information provided by pair (r^a, r^b). Assuming each cluster as a Gaussian distribution, we use the Wasserstein distance [20] to measure the distance of every two clusters c^a and c^b as

$$-d_{inter}^2(c^a, c^b) = -\mathcal{W}_2(\mathcal{N}(f_{mean}^a, \Sigma^a); \mathcal{N}(f_{mean}^b, \Sigma^b))^2 =$$
$$\|f_{mean}^a - f_{mean}^b\|_2^2 + Tr(\Sigma^a + \Sigma^b - 2(\Sigma^{a1/2}\Sigma^b\Sigma^{a1/2})^{1/2}) \quad (6)$$

where the mean f_{mean}^a and the variance Σ^a come from embedding f_i^a of all images $x_i \in c^a$, and every dimensions of f are considered to be independent from each other.

3.4.3. Cluster re-assignment

In the intra-cluster stage, once annotators observed $v(r^a, x_i^a) = 0$, i.e. $y_{r^a} \neq y_{x_i^a}$, we separate them into two clusters and divide rest samples in c^a according to their distance. In the inter-cluster stage, if $v(r^a, r^b) = 1$ i.e. $y_{r^a} = y_{r^b}$, we set $\widetilde{y}^a = \widetilde{y}^b = \widetilde{y}_{r^j}^a$.

3.5. Diversity Criterion

Considering one extreme situation that all selected pairs are relevant to one representative sample r^a, as the information provided by r^a is limited, this would result in the redundancy in $\mathbb{P}_{sel}$. Therefore, we also strive for the diversity of selected pairs. Intuitively, we expect that when annotating, annotators have an equal probability of concerning each cluster. Therefore, we measure the diversity as the KL-divergence between the frequency of each cluster in selected pair set $\mathbb{P}_{sel}$ and the uniform distribution, termed as the diversity criterion for pair selection. Let p_j denotes the probability of c^j in $\mathbb{P}_{sel}$ as

$$p_j = \frac{cnt(c^j)}{T \times 2} \quad (7)$$

where $cnt(c^j)$ counts the times that $x^j \in c^j$ shows in $\mathbb{P}_{sel}$. So the KL-divergence between p_j and the uniform distribution $U(0,1)$ is measured as

$$D_{KL}(Q\|P) = \sum_{j=1}^{|C|} q \times (\log\frac{q}{p_j}) \quad (8)$$

where Q is a uniform distribution.

3.6. Solving Stage Optimization Objective

So far, we've simplified and specified Ep. 1. The final optimization objective function is written as

$$\min O(\mathbb{P}_{sel}) = D + \alpha D_{KL} = \Sigma d + \alpha D_{KL}$$
$$s.t. |\mathbb{P}_{sel}| = T, \mathbb{P}_{sel} \in \mathbb{P}_{stage} \quad (9)$$

where D denotes the specific distance measuring for each stage, i.e., for the intra-cluster stage, D means the chaotic distance, while for the inter-cluster stage, D means the Wasserstein distance. To solve this objective, however, it is hard to get the accurate optimal solution.

Therefore, we use a greedy strategy to obtain a near-optimum solution, which selects a pair and pass to the human annotator once a time, according to which pair minimizes the increment of O the most at each step, until the upper bound budget of annotation budget T is reached. For each pair (x^a, x^b), the increment of O if this pair is selected is represented as

$$o_{inc}(x^a, x^b) = d(x^a, x^b) + \frac{\alpha}{T}o_{inc:D(KL)(x^a, x^b)} \quad (10)$$

Because q is a constant as we fix T,
$o_{inc:D(KL)}$ is only related to the change of $\log cnt(c^j)$ at each step. For example, if pair(x^a, x^b) is selected at step t, then we have

$$o_{inc:D(KL)}(x^a, x^b) = \beta[\log cnt_{t-1}(c^a) - \log cnt_t(c^a)$$
$$+ \log cnt_{t-1}(c^b) - \log cnt_t(c^b)] \quad (11)$$

where β is also a constant calculated by q, α and T. The $cnt(c^a)$ will increase by 1 or 2 once the pair that includes samples from c^a is selected at each step t. Because of the diversity criterion we use, selections are relatively sparse, so one pair has limited effect on others, and the greedy strategy is sufficient to get a near-optimum solution.

3.7. Iterative Model Training

By iterative executing the clustering-selecting-annotation-reassigning-training procedure for iterations, $F(x;\theta)$ will be finally obtained. We distributed the annotation budget to each epoch in a Gaussian manner, with the mean and variance set to epoches/2. So there are more annotation times assigned in the middle of training. This is because at the beginning, the discriminability of the model $F(x;\theta)$ is limited and valuable pairs are hard to select. On the other hand, at the end, $F(x;\theta)$ has become stable and annotations cannot be fully utilized.

4. EXPERIMENTS

4.1. Datasets and Evaluation Protocols

The Market-1501 [17] dataset contains 32668 images of 1501 identities. DukeMTMC-ReID [18], which is a subset of DukeMTMC [23], contains 36411 images of 1404 identities in total. The MSMT17[24] dataset contains 4101 identities with 126,441 images. We use the Cumulative Matching Characteristic (CMC) curve and the mean average precision (mAP) to evaluate the proposed method.

	Method	Market1501				DukeMTMC-reID			
		mAP	Rank-1	Rank-5	Rank-10	mAP	Rank-1	Rank-5	Rank-10
Un	HCT [9]	56.4	80.0	91.6	95.2	50.7	69.6	83.4	87.4
	MMCL [10]	45.5	80.3	89.4	92.3	40.2	65.2	75.9	80.0
	SPCL(Un) [14]	73.1	88.1	95.1	97.0	-	-	-	-
	C-contrast/ResNet50 [15]	82.6	93.0	97.0	98.1	72.8	85.7	92.0	93.5
UDA	AD-C [12]	68.3	86.7	94.4	96.5	54.1	72.6	82.5	85.5
	MMT [13]	71.2	87.7	94.9	96.9	65.1	78.0	88.8	**92.5**
	JVTC+ [21]	67.2	86.8	95.2	97.1	66.5	80.4	89.9	92.2
Semi	SGCDPL [1]	76.4	91.1	-	-	66.5	82.2	-	-
	MTML [22]	65.2	85.3	96.2	97.6	50.7	71.7	86.9	89.6
Active	DRAL [4]	66.3	84.2	94.3	96.6	56.0	74.3	84.8	88.4
	MASS [5]	81.7	93.5	97.9	98.6	72.9	86.1	93.9	95.9
Active	Ours-baseline-GEMpooling	81.6	91.6	96.7	97.9	70.1	83.2	91.0	93.2
	Ours-Maxpooling	82.8	92.5	97.2	98.0	74.5	86.4	92.8	94.7
	Ours-GEMpooling	**85.6**	**93.6**	**97.7**	**98.4**	**75.3**	**86.9**	**92.8**	**94.7**

Table 1. Compare with state-of-the-art method. The comparison of the annotation cost is included in section 4.5.

Method	mAP	Rank-1	Rank-5	Rank-10
(active) MASS [5]	30.0	54.1	65.4	70.4
(Un) C-Contrast [15]	33.3	63.3	73.7	77.8
Ours-baseline	30.7	59.0	70.7	75.3
Ours	**35.5**	**63.4**	**74.5**	**79.0**

Table 2. Compare with state-of-the-art method on MSMT17.

Method	mAP	Rank-1	Rank-5	Rank-10
Ours w/o intra.	83.2	93.4	97.4	98.5
Ours w/o inter.	82.5	92.1	96.9	98.0
Ours w/o diver.	84.9	93.5	97.6	98.5
Ours	85.6	93.6	97.7	98.4
Ours-upper	87.5	94.6	97.7	98.7

Table 3. Ablation studies: methods with different modules and the fully-supervised upper bounds on Market1501. Our method is close to the upper bound.

Fig. 2. Visualization of p_j with (right) and without (left) the diversity criterion.

method	T	mAP	Rank-1	Rank-5
DRAL [4]	$10N \approx 120k$	66.3	84.2	94.3
MASS [5]	$5N \approx 60k$	81.7	93.5	97.9
Ours	1000	84.1	93.3	97.4
	3000	**85.6**	**93.6**	**97.7**

Table 4. mAP and Rank-1 with different T on Market1501. We can greatly reduce the annotation cost while achieving better performance.

4.2. Implementation Details

ResNet-50 [25] is adopted as the CNN model followed by a GEM pooling [26] and a BatchNorm [27] layer. The input images will be transformed with flipping and random erasing [28]. We use Adam [29] as the optimizer with weight decay 0.0005. The learning rate is set to 0.00035. Following the baseline method [15], DBSCAN [30] is chosen with ϵ set to 0.4/0.7 and the minimum number of points required to form a dense region is set to 4 with image Jaccard distance [31].

4.3. Results and Comparisons with State-of-the-Arts

Results show in Table 1 and 2, with T set to 3000. The baseline method does not need annotations. Unsupervised methods are marked as 'Un', unsupervised person re-id methods are marked as 'UDA'. Our method can outperform them all with simpler structure design. Compared with semi-supervised methods. Our proposal also shows good performance even with less annotations. Compare with the active learning based method DRAL [4, 5], our method also outperforms them, with much less annotation cost.
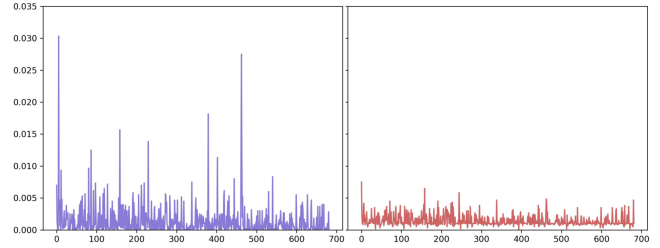
4.4. Ablation Studies

4.4.1. Ablation of modules

The ablation study is in Table 3. 'Ours w/o intra' and 'Ours w/o inter' show separately the Inter/Intra pair selecting and pseudo label updating strategy both improve the performance. Moreover, by replacing pseudo labels with ground-truth labels, we report the upper bound performance of our method, which trains the model with full supervision. We can see that **the performance of our model is quite close to 'Ours-upper'**, which verifies our aim: reducing the annotation cost while maintaining the performance. We also discuss the effect of the diversity regularized selection through Figure 2. It shows the values of p_j defined from Eq. 7 of each representative sample r^j. Through our KL-divergence regularized criterion, p_j will be more uniform, which means the annotator looks at each sample with a relatively uniform probability.

4.5. Effects of Annotates and Annotate Costs

Table 4 shows the performance with different budget T. We can see that with a larger T, the model performs better, which is consistent with the intuition that more supervision leads to better results. The total annotation cost of our proposal in all epochs T is set to 3000. We compare our annotation cost with MASS [5] and DRAL [4], which is the state-of-the-art methods for reducing the annotation cost for the Re-ID task. $N \times 5$ in [5] is reported to be the annotation cost, and $N \times 10$ in [5] is the adopted budget. Our budget is $T < N < N \times 5 < N \times 10$, meaning that we use a much lower annotation cost to achieve better results.

5. CONCLUSION

To reduce the annotation cost while maintaining the performance for the Re-ID task, in this paper we propose the Annotation Efficient Person Re-Identification method to select image pairs from an alternative pair set according to the fallibility and diversity of pairs, and train the Re-ID model based on the annotation. Extensive experiments show that we can achieve better performance with much less annotations compared with current works.

6. REFERENCES

[1] Chih-Ting Liu, Yu-Jhe Li, and Shao-Yi Chien, et al., "Semantics-guided clustering with deep progressive learning for semi-supervised person re-identification," *arXiv preprint arXiv:2010.01148*, 2020.

[2] Xiaomeng Xin, Jinjun Wang, and Ruji Xie, et al., "Semi-supervised person re-identification using multi-view clustering," *PR*, vol. 88, pp. 285–297, 2019.

[3] Xiaomeng Xin, Xindi Wu, and Yuechen Wang, et al., "Deep self-paced learning for semi-supervised person re-identification using multi-view self-paced clustering," in *ICIP*. IEEE, 2019, pp. 2631–2635.

[4] Zimo Liu, Jingya Wang, and Shaogang Gong, et al., "Deep reinforcement active learning for human-in-the-loop person re-identification," in *ICCV*, 2019, pp. 6122–6131.

[5] Bingyu Hu, Zheng-Jun Zha, and Jiawei Liu, et al., "Cluster and scatter: A multi-grained active semi-supervised learning framework for scalable person re-identification," in *ACM MM*, 2021, pp. 2605–2614.

[6] Jian Wu, Victor S Sheng, and Jing Zhang et al., "Multi-label active learning algorithms for image classification: Overview and future promise," *CSUR*, vol. 53, no. 2, pp. 1–35, 2020.

[7] Hanxiao Wang, Shaogang Gong, Xiatian Zhu, and Tao Xiang, "Human-in-the-loop person re-identification," 2018.

[8] Xin Xu, Lei Liu, and Xiaolong Zhang, et al., "Rethinking data collection for person re-identification: active redundancy reduction," *PR*, vol. 113, pp. 107827, 2021.

[9] Kaiwei Zeng, Munan Ning, and Yaohua Wang, et al., "Hierarchical clustering with hard-batch triplet loss for person re-identification," in *CVPR*, 2020, pp. 13657–13665.

[10] Dongkai Wang and Shiliang Zhang, "Unsupervised person re-identification via multi-label classification," in *CVPR*, 2020, pp. 10981–10990.

[11] Jinlin Wu, Yang Yang, and Hao Liu, et al., "Unsupervised graph association for person re-identification," in *ICCV*, 2019, pp. 8321–8330.

[12] Yunpeng Zhai, Shijian Lu, and Qixiang Ye, et al., "Ad-cluster: Augmented discriminative clustering for domain adaptive person re-identification," in *CVPR*, 2020, pp. 9021–9030.

[13] Yixiao Ge, Dapeng Chen, and Hongsheng Li, "Mutual mean-teaching: Pseudo label refinery for unsupervised domain adaptation on person re-identification," *arXiv preprint arXiv:2001.01526*, 2020.

[14] Yixiao Ge, Dapeng Chen, and Feng Zhu, et al., "Self-paced contrastive learning with hybrid memory for domain adaptive object re-id," *arXiv preprint arXiv:2006.02713*, 2020.

[15] Zuozhuo Dai, Guangyuan Wang, and Weihao Yuan, et al., "Cluster contrast for unsupervised person re-identification," *arXiv preprint arXiv:2103.11568*, 2021.

[16] Kaiwei Zeng, Munan Ning, and Yaohua Wang, et al., "Energy clustering for unsupervised person re-identification," *IVC*, p. 103913, 2020.

[17] Liang Zheng, Liyue Shen, and Lu Tian, et al., "Scalable person re-identification: A benchmark," in *ICCV*, December 2015.

[18] Zhedong Zheng, Liang Zheng, and Yi Yang, "Unlabeled samples generated by gan improve the person re-identification baseline in vitro," in *ICCV*, Oct 2017.

[19] Wenhe Liu, Xiaojun Chang, and Ling Chen et al., "Pair-based uncertainty and diversity promoting early active learning for person re-identification," *TIST*, vol. 11, no. 2, pp. 1–15, 2020.

[20] Ingram Olkin and Friedrich Pukelsheim, "The distance between two random vectors with given dispersion matrices," *LinAlgApp*, vol. 48, pp. 257–263, 1982.

[21] Jianing Li and Shiliang Zhang, "Joint visual and temporal consistency for unsupervised domain adaptive person re-identification," in *ECCV*. Springer, 2020, pp. 483–499.

[22] Xiangping Zhu, Xiatian Zhu, and Minxian Li, et al., "Intra-camera supervised person re-identification: A new benchmark," in *ICCVW*, 2019, pp. 0–0.

[23] Ergys Ristani, Francesco Solera, and Roger Zou, et al., "Performance measures and a data set for multi-target, multi-camera tracking," in *ECCV*. Springer, 2016, pp. 17–35.

[24] Longhui Wei, Shiliang Zhang, and Wen Gao, et al., "Person transfer gan to bridge domain gap for person re-identification," in *CVPR*, 2018, pp. 79–88.

[25] Kaiming He, Xiangyu Zhang, and Shaoqing Ren, et al., "Deep residual learning for image recognition," in *CVPR*, 2016, pp. 770–778.

[26] Yinzheng Gu, Chuanpeng Li, and Jinbin Xie, "Attention-aware generalized mean pooling for image retrieval," *arXiv preprint arXiv:1811.00202*, 2018.

[27] Sergey Ioffe and Christian Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.

[28] Zhun Zhong, Liang Zheng, and Guoliang Kang et al., "Random erasing data augmentation," in *AAAI*, 2020, vol. 34, pp. 13001–13008.

[29] Diederik P Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[30] Martin Ester, Hans-Peter Kriegel, and Jörg Sander, et al., "A density-based algorithm for discovering clusters in large spatial databases with noise.," in *KDD*, 1996, vol. 96, pp. 226–231.

[31] Zhun Zhong, Liang Zheng, and Donglin Cao, et al., "Re-ranking person re-identification with k-reciprocal encoding," in *CVPR*, 2017, pp. 1318–1327.