

Mitigate Catastrophic Remembering via Continual Self-Paced Dual-Knowledge Purification for Noisy Lifelong Person Re-Identification

Kunlun Xu¹, Jiangmeng Li¹, Yufei Guo¹, Yuxin Peng¹, *Fellow, IEEE*, and Jiahuan Zhou¹

Abstract—Current Lifelong Person Re-Identification (LReID) methods focus on tackling the data streams with accurate labels. When the data streams with noisy labels are given, their performance is severely degraded since the model continually remembers erroneous knowledge induced by the label noise. Moreover, the well-known issue of catastrophic forgetting in LReID is exacerbated by noisy labels, which disrupt the retention of correct knowledge from previous models. Such a practical Noisy LReID task is important but challenging, and few works have attempted to address it. In this paper, we pioneer the investigation of Noisy LReID and propose a Continual Self-Paced Dual-Knowledge Purification (CSDP) method to address the catastrophic remembering of erroneous knowledge, catastrophic forgetting of correct knowledge, and limited acquisition capacity, simultaneously. Specifically, both local distribution information and global classification results are jointly leveraged to enhance the recognition and correction of noisy labels. In addition, a self-paced learning paradigm is introduced to facilitate the progressive accumulation of reliable knowledge. Finally, a Noisy LReID benchmark is constructed to support research on this task. Extensive experimental results demonstrate that the proposed CSDP method, under comparable computational overhead, outperforms state-of-the-art methods by a large margin.

Index Terms—Lifelong learning, person re-identification, label noise learning, knowledge purification.

I. INTRODUCTION

PERSON re-identification (ReID) [2], [3], [4] aims to match images of the same individual across different non-overlapping camera views. Most existing ReID studies focus

on stationary settings, where the camera environments remain unchanged [5], [6], [7]. Recently, lifelong person re-identification (LReID) [8], [9], [10], [11] has attracted increasing attention, as it aims to learn continuously from non-stationary data streams, resembling the human learning process over time. Despite this progress, most existing LReID methods assume perfectly annotated training data [1], [12], [13]. In practice, labels are often noisy due to inaccurate person detection or annotation errors [14], [15], and prior studies have shown that such noise can significantly degrade ReID performance in stationary scenarios [16], [17].

In the LReID scenario, models are particularly vulnerable to label noise. As illustrated in Fig. 1(a), traditional methods are trained on clean data with correct labels, thereby learning accurate affinity relationships. In contrast, in noisy data, label noise (*i.e.*, incorrect identity annotations) introduces erroneous affinity. As shown in Fig. 1(b), when trained on noisy data, the model learns incorrect identity affinities [17], resulting in noticeable performance degradation on new data. Furthermore, as illustrated in Fig. 1(c), even if correct knowledge has been acquired from earlier stages, label noise in subsequent data can inject erroneous information, causing the model to forget previously learned correct affinity [18], [19]. This effect exacerbates catastrophic forgetting [20] and leads to degraded performance on historical data. In addition, as shown in Fig. 1(d), the model may accumulate erroneous affinity learned from earlier noisy stages, which further hinders the acquisition of correct knowledge in later stages, ultimately degrading the average test performance. This phenomenon is denoted as catastrophic remembering. These issues substantially impair the effectiveness of existing LReID and continual learning methods [1], [9], [11], [12], [21].

One straightforward solution for the Noisy LReID problem is to incorporate the label noise learning (LNL) [14], [22] approaches with the LReID methods. In LNL, filtering- and rectification-based methods have demonstrated advanced performance, particularly in ReID [17]. Most existing approaches rely on Gaussian Mixture Models (GMM) [3], [17], [23] to detect incorrect labels or adopt k-Means clustering [24], [25], [26] to relabel noisy samples. Despite their success, as shown in Fig. 2(a), GMM-based methods rely on classification loss and often discard correct samples that deviate from the class center. Meanwhile, as shown in Fig. 2(b), cluster-based methods enforce a fixed number of clusters. However, in ReID, a single

Received 26 September 2025; revised 18 April 2026; accepted 9 May 2026. Date of publication 15 May 2026; date of current version 4 September 2026. This work was supported in part by the National Key R&D Program of China under Grant 2024YFA1410000, in part by the National Natural Science Foundation of China under Grant 62376011, Grant 62525201, Grant 62132001, Grant 62432001, and Grant 62406313, in part by Beijing Natural Science Foundation under Grant L247006 and Grant L257005, and in part by the Postdoctoral Fellowship Program of China Postdoctoral Science Foundation under Grant YJB20250283. Recommended for acceptance by W. Zheng. (Kunlun Xu and Jiangmeng Li contributed equally to this work.) (Corresponding author: Jiahuan Zhou.)

Kunlun Xu, Yuxin Peng, and Jiahuan Zhou are with the Wangxuan Institute of Computer Technology, Peking University, Beijing 100080, China (e-mail: xkl@stu.pku.edu.cn; pengyuxin@pku.edu.cn; jiahuanzhou@pku.edu.cn).

Jiangmeng Li is with the Institute of Software, Chinese Academy of Sciences, Beijing 100190, China (e-mail: jiangmeng2019@iscas.ac.cn).

Yufei Guo is with Intelligent Science & Technology Academy of CASIC, Beijing 100041, China (e-mail: yfguo@pku.edu.cn).

Our code is available at <https://github.com/zhouljiahuan1991/TPAMI-CSDP>. This article has supplementary downloadable material available at <https://doi.org/10.1109/TPAMI.2026.3693671>, provided by the authors.

Digital Object Identifier 10.1109/TPAMI.2026.3693671

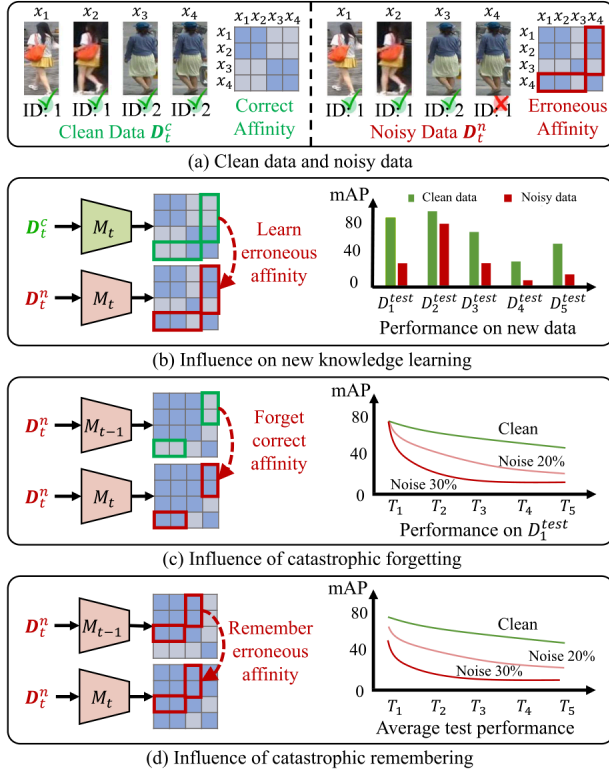


Fig. 1. (a) Visualization of the labeled data and the corresponding affinity matrix, which represents the pairwise matching scores between instances. Noisy data contains wrong labels that introduce erroneous inter-instance affinity knowledge. During LReID, noisy data not only (b) influences the knowledge learning of new datasets but also (c) exacerbates the forgetting of correct historical knowledge. Besides, (d) the learned erroneous knowledge is remembered by the new model M_t which severely hinders the LReID performance. The experiment results are obtained from the latest LReID method [1].

identity often exhibits substantial appearance variation due to changes in viewpoint, pose, and background, while different identities may share high similarity in these visual characteristics [27], [28]. Therefore, these methods tend to mix samples from different identities and generate incorrect labels, which can constrain model performance. Consequently, when existing LNL methods are directly applied to the Noisy LReID scenario, the challenges illustrated in Fig. 1(b), (c), and (d) persist.

In our conference version, we proposed CKP (Fig. 2(c)), which exploits local relations to identify noisy labels by allowing multiple adaptive local clusters within each identity, thereby capturing intra-identity diversity. However, CKP still faces two challenges: 1) Inevitable noise within local clusters and neglected outliers during clustering, leading to limited data utilization capacity. 2) After each data purification iteration, newly identified hard samples are incorporated together with easy samples for training. However, since these hard samples often exhibit features that differ significantly from the current model's learned representations, directly optimizing on them may corrupt the previously learned robust features and lead the model to suboptimal solutions [29]. This, in turn, intensifies catastrophic forgetting and disrupts the overall learning process.

To address the above challenges, we propose a novel method named Continual Self-Paced Dual-Knowledge

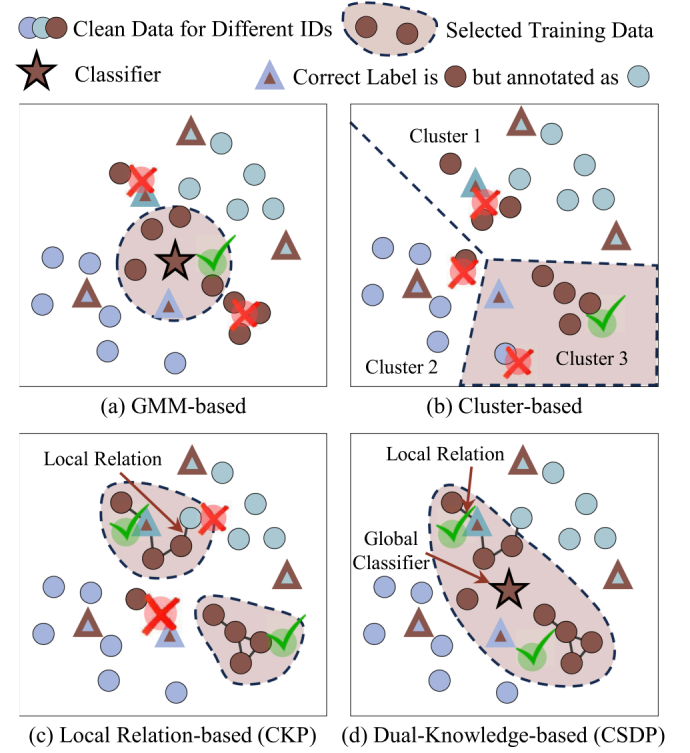


Fig. 2. Comparison of different noisy label filtering and rectification methods.

Purification (CSDP). Specifically, we design a Local-Global Collaborative Purification (LGCP) mechanism that jointly leverages the global knowledge within classification scores and local relations to discard noisy labels while retaining as many correct ones as possible, as illustrated in Fig. 2(d). Then, to improve new knowledge acquisition and handle catastrophic remembering (Fig. 1(d)) during model training, we introduce a Self-Paced Noisy Knowledge Evolution scheme (SNKE), where the purified data are divided into subsets of increasing hardness, upon which a self-paced learning mechanism is built to progressively incorporate harder samples to enhance model learning. In addition, an Erroneous Knowledge Filtering algorithm (EKF) is proposed, which estimates the correctness of historical model outputs and selectively transfers correct knowledge while excluding erroneous ones. Moreover, instead of permanently dropping mislabeled samples, we design an Iterative Label Rectification (ILR) scheme that continuously refines their labels during training. These rectified samples can be reintroduced into LGCP for reuse, thereby improving both data efficiency and model robustness. Therefore, the above modules work collaboratively to continually purify the new and historical knowledge under the lifelong noisy learning scenario. Finally, we construct a new Noisy LReID Benchmark (NLReID), which contains 5 training domains and 12 testing domains. It includes 2,500 training identities and over 40,000 images in total. Experimental results demonstrate that CSDP outperforms state-of-the-art methods across diverse noisy conditions. In particular, CSDP achieves over 10.0% improvement when the proportion of noisy labels reaches 30%.

In summary, our contributions are fourfold: 1) We pioneer the study of the Noisy LReID problem and construct a comprehensive Noisy LReID benchmark (NLReID) to facilitate evaluation. In addition, we propose a novel CSDP framework to address the critical issues of catastrophic forgetting and remembering under label noise. 2) A Local-Global Collaborative Purification mechanism is developed to jointly exploit complementary knowledge from classification scores and local relations, enabling more accurate new knowledge learning while mitigating the acquisition of erroneous knowledge. 3) A Self-Paced Noisy Knowledge Evolution scheme is introduced, which incorporates self-paced noisy model optimization and erroneous knowledge filtering algorithms to improve learning stability and mitigate catastrophic remembering. 4) Extensive experiments under diverse noise settings demonstrate the effectiveness and superiority of the proposed CSDP framework compared to state-of-the-art methods.

This manuscript extends our previous conference paper [30] with substantial improvements in methodology, analysis, and experiments: 1) To address the limitations of noisy clustering and overlooked outliers, we propose a Local-Global Collaborative Purification mechanism that jointly exploits complementary knowledge from classification scores and local relations, thereby enhancing correct knowledge acquisition and mitigating the influence of erroneous labels. 2) To alleviate model instability caused by directly utilizing all purified data, we introduce a Self-Paced Noisy Knowledge Evolution scheme that progressively incorporates samples of increasing difficulty, improving model robustness and learning capacity. 3) Comprehensive theoretical justifications are provided in Section IV to validate the effectiveness and rationality of our method. 4) More ablation studies and detailed analyses, including ablation studies on module effectiveness and ablation studies on different baselines. 5) More quantitative and visualization results on the model learning process, including data purification results, model learning trends, and training/testing feature distributions as presented in Section V.

II. RELATED WORK

A. Lifelong Person Re-Identification

Lifelong person re-identification (LReID) [8], [9], [10], [31], [32] aims to continuously train a ReID model on non-stationary data, enhancing its adaptability to diverse environments. Several research groups [1], [9], [14] have made significant contributions to the LReID field, forming a solid foundation for this line of research. A major challenge in LReID is catastrophic forgetting [33], where performance on previously learned domains deteriorates sharply when adapting to new ones [32], [34], [35]. A straightforward solution is to store and replay historical exemplars [8], [36]. However, preserving human images raises privacy concerns and increases storage costs. Consequently, existing LReID methods primarily pursue exemplar-free strategies, such as knowledge distillation, prototype-based learning, and style transfer. For instance, LSTKC [1] and PatchKD [11] employ distillation losses to preserve historical representations, while DKP [12] leverages distributional prototypes for better

historical information retention. DASK [37] further introduces an input-level rehearsal mechanism that transforms new data into historical styles, thereby mitigating forgetting.

Despite these advances, prior works [8], [9], [13], [37], [38], [39] generally assume access to perfectly annotated training data. In real-world scenarios, labels are often noisy due to imperfect person detection or annotation errors [14], [15], [17], [40], [41]. This *Noisy LReID* setting introduces additional challenges: i) Noisy labels inject incorrect information into the learning process. They can overwrite previously learned correct knowledge and exacerbate catastrophic forgetting. ii) The accumulation of erroneous knowledge across historical domains may impair the model's generalization capacity, leading to the phenomenon of catastrophic remembering [42], [43]. iii) The noisy labels can disturb the acquisition of correct new knowledge, resulting in reduced learning capacity. Furthermore, under a higher noise ratio, these methods suffer from increased performance degradation. In this work, we present a pioneering systematic investigation of the Noisy LReID problem and propose an effective baseline to address this challenging setting.

B. Label Noise Learning

Label Noise Learning (LNL) aims to improve the robustness of models against noisy annotations [44], [45], [46], [47], [48], [49]. Existing approaches can be broadly categorized into data modification [23], [50] and loss regularization [51], [52], [53].

Data modification methods attempt to filter or rectify noisy data to enhance correct knowledge learning. Most of these methods rely on the small-loss trick to identify clean samples. For example, LCNL [17] and TCL [23] adopt Gaussian Mixture Models (GMM) to filter noisy data, while DivideMix [54] integrates GMM with semi-supervised learning (e.g., MixMatch [55]) to relabel noisy data. Although effective under moderate noise, their performance deteriorates under high noise due to error accumulation in pseudo-labeling. Recently, clustering-based strategies have been explored. PNAL [48] and STC [25] rectify labels using clustering results, and HOC [56] estimates the label noise transition matrix based on clusterability. Benefiting from the robustness of cluster parameters and extracted local relations, these methods achieved improved noise recognition and correction capacity [56], [57] in the classification task. However, person re-identification is a fine-grained, instance-level task, suffering from large intra-identity diversity [58], [59]. In this setting, clustering results are often noisy themselves, making it unreliable to depend solely on clustering.

Loss regularization methods instead modify the training objective to reduce noise sensitivity. Forward and backward correction strategies leverage noise transition matrices to maintain statistical consistency [60], [61], while methods such as CORE [14] introduce regularization terms to suppress label bias in high-confidence predictions [62], [63]. However, these approaches remain vulnerable to memorization effects when fitting hard-noise samples. Moreover, noisy samples near distribution boundaries often exhibit regularization effects similar to clean samples, leading to weakened noise robustness [30].

TABLE I
SUMMARY OF THE KEY NOTATIONS IN THIS PAPER

Notation	Description
e	The index of model training epoch
r	The index of framework cycle round
t	The index of the current training step
T	The number of training datasets
d_i	The label distance of an instance x_i
s_i	The data confidence score of an instance x_i
s_i^l	The label confidence score of an instance x_i
w_i^o	The knowledge distillation weight
D_t	The current training dataset
D_t^r	The updated training dataset after r -th round
D_t^{r*}	The clean subset of D_t^r
M_t^r	The current model trained after r -th round
M_{t-1}	The final model obtain by the previous training step

Motivated by these limitations, we propose to exploit local relations to capture fine-grained cues under large inter-identity diversity. Furthermore, we integrate model predictions with local relations to improve noisy-label filtering and maximize the utilization of clean data.

C. Lifelong Learning With Label Noise

The problem of Lifelong Learning with Label Noise has received limited attention, and existing research mainly focuses on classification tasks [64], [65], [66], [67]. Most methods rely on filtering and storing historical exemplars to alleviate catastrophic forgetting. For example, SPR [64] observed that replaying noisy data may aggravate forgetting and thus designed a self-purified replay strategy that stores only clean samples. Similarly, CNLL [65] maintains separate replay buffers for clean and noisy data, enabling a joint utilization of both reliable and unreliable data.

However, in practice, human images are highly privacy-sensitive [12], making exemplar retention infeasible in many real-world applications [9], [11]. This limitation renders existing exemplar-based solutions impractical for lifelong person re-identification [1], [37]. Besides, these methods typically require linearly increasing storage and training overhead over time. This further limits their application in real scenes.

To address this gap, we present a pioneer systematic study of *Noisy LReID*, where catastrophic forgetting of correct knowledge and catastrophic remembering of erroneous knowledge are simultaneously investigated in an exemplar-free manner.

III. THE PROPOSED METHOD

A. Problem Definition and Formulation

The key notations of this paper are summarized in Table I. Specifically, noisy lifelong person re-identification (Noisy LReID) aims to continuously learn from a stream of T ReID datasets $\mathcal{D}^{tr} = \{D_t\}_{t=1}^T$, where each dataset contains a certain proportion of noisy labels. The final model is evaluated on the corresponding clean test sets $\mathcal{D}^{te} = \{D_t^{te}\}_{t=1}^T$ to measure both new knowledge acquisition and anti-forgetting ability. In addition, a set of U extra clean test sets $\mathcal{D}^{un} = \{D_t^{un}\}_{t=1}^U$ is used to assess the model's generalization to unseen domains.

In this paper, the obtained model by training on dataset D_t is denoted as M_t . Moreover, the training process on each dataset is divided into a set of rounds, where the intermediate model after the r -th round on D_t is denoted as M_t^r .

B. Overview

As shown in Fig. 3, given a noisy training dataset D_t at training step t , our approach iteratively generates a rectified dataset D_t^r at training round r , where D_t^0 is initialized as the original noisy dataset D_t . The proposed Continual Self-Paced Dual-Knowledge Purification (CSDP) framework consists of three key components: Local-Global Collaborative Purification (LGCP), Self-Paced Noisy Knowledge Evolution (SNKE), and Iterative Label Rectification (ILR). Specifically, LGCP uses two sources of information: local relationships between instances and global classification results. Based on these, it estimates how confident the label of each instance is. Then it builds a clean and reliable subset, denoted as D_t^{r*} . This subset helps guide the model to learn new data accurately. Building upon this, SNKE introduces a self-paced learning mechanism that gradually incorporates hard samples, enhancing model learning stability and robustness. In addition, an Erroneous Knowledge Filtering mechanism within SNKE evaluates the correctness of knowledge of previous models, enabling the active forgetting of erroneous knowledge while consolidating correct knowledge. Finally, at the end of the r -th round, the ILR scheme rectifies noisy labels by integrating model predictions with original label information, producing an updated rectified dataset D_t^{r+1} for subsequent training.

C. Local-Global Collaborative Purification

Given the noisy dataset D_t , we iteratively update the labels of the images as introduced in Section III-E. The updated dataset after r -th round is denoted as D_t^r . As is shown in Fig. 3, at the r -th round, D_t^{r-1} and M_t^{r-1} is used to process the images and obtain the features $\mathcal{F}_t^r = \{f_i^r\}_{i=1}^{N_t}$. Then, to utilize the complementary information between local distribution and global view, we introduce a local-global collaborative paradigm to purify the noisy labels. Specifically, a Cluster-aware Label Scoring (CLS) module and a Dual-Knowledge Guided Purification mechanism are exploited as introduced in the following.

1) *Cluster-Aware Label Scoring*: To mine the local relation across instances, we propose to utilize the density-based clustering [68] to adaptively gather instances with shared characteristics, based on which to achieve label confidence estimation. Specifically, given the noisy dataset $\{(x_i, y_i)\}_{i=1}^{N_t}$ with N_t images x_i and corresponding labels y_i , the extracted features are $\{f_i\}_{i=1}^{N_t}$. Then, the density-based clustering algorithm (e.g., DBSCAN [68]) is adopted to generate N_c non-overlapping clusters, where N_c is automatically determined. Each instance is assigned a cluster-aware label according to its cluster, i.e. $\tilde{y}_i \in \{1, 2, \dots, N_c\}$. Note that the outliers during the clustering process are collected as an extra cluster whose cluster label is set to $N_c + 1$. Then, we generate one-hot embedding $\mathbf{l}_i \in \mathbb{R}^{N_c}$ for all instances.

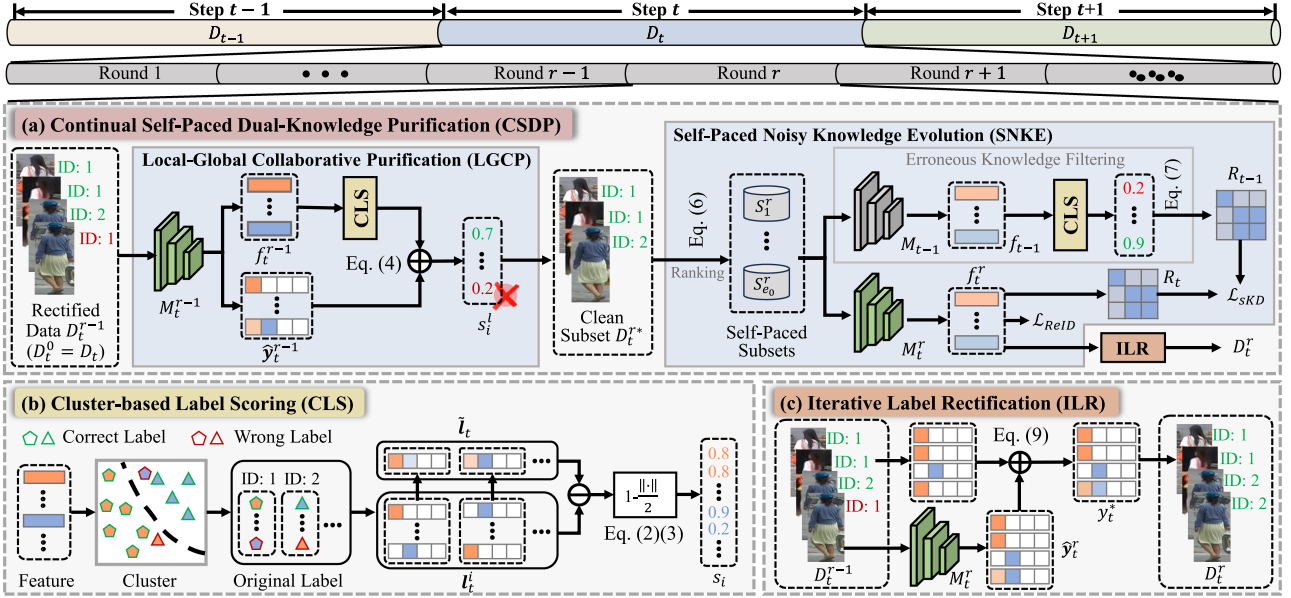


Fig. 3. Given the noisy dataset D_t , our CSDP updates the rectified dataset D_t^r and model M_t^r along the module process rounds r . The LGCP module aims to obtain a clean subset D_t^{r*} for M_t^r learning. Besides, the SNKE module introduces a self-paced learning paradigm to motivate robust learning within each round and exploits an Erroneous Knowledge Filtering mechanism to filter the features containing erroneous old knowledge, mitigating catastrophic remembering. Finally, ILR rectifies the noise labels by fusing the learned knowledge in M_t^r and label information to further improve the utilization of noisy data.

Then, to bridge the connection between the annotated label and the generated cluster label, an annotation-aware average cluster label $\bar{l}_i$ for each instance x_i is obtained by

$$\bar{l}_i = \frac{1}{n_t^i} \sum_{j=1}^{N_t} \delta(y_i, y_j) l_j, \quad (1)$$

where $n_t^i = \sum_{j=1}^{N_t} \delta(y_i, y_j)$, and $\delta(y_i, y_j)$ is a sign function that outputs 1 and 0 when $y_i = y_j$ and $y_i \neq y_j$ respectively. Equation (1) indicates that given x_i with annotated label y_i , obtain the average cluster label of all instances x_j that possess the same annotated label as y_i . Therefore, $\bar{l}_i$ is consistent across instances with the same annotated label, representing the aggregate value of each annotated label in the cluster label space.

To quantify the confidence of each instance x_i , the label distance d_i is defined as the squared L2 norm between l_i and $\bar{l}_i$, represented as:

$$d_i = \|l_i - \bar{l}_i\|_2^2, \quad (2)$$

where d_i measures the disparity between the cluster label $\tilde{y}_i$ and the annotated identity centers. Note that $d_i \in [0, 2)$ (the justification of this range is provided in Section IV-A). Besides, the d_i values of outliers during clustering are adjusted to 2. This adjustment is made because the outliers exhibit minimal resemblance to other instances and thus possess the lowest confidence. Then the data confidence score s_i is calculated by

$$s_i = (2 - d_i)/2, \quad (3)$$

where $s_i \in [0, 1]$ with higher values indicating greater trustworthiness of the annotation.

2) *Dual-Knowledge Guided Purification*: Relying solely on local relations can be misguided by the noisy clustering results

and neglect the utilization of outlier samples, as shown in Fig. 2(c). To overcome this issue, we jointly utilize the local relation score s_i extracted by CLS and the classification results that contain the identity-level global clue. To achieve this, we generate the label confidence score s_i^l of each instance x_i by

$$s_i^l = s_i * w_c + \hat{y}_t^{r-1}[y_i] * (1 - w_c), \quad (4)$$

where w_c is a hyperparameter to balance the weight of local and global knowledge and $\hat{y}_t^{r-1}$ is the identity-wise confidence vector generated by M_t^{r-1} .

Then, we remove the data of low-confidence labels with a confidence threshold T_c and obtain the clean subset D_t^{r*} :

$$D_t^{r*} = \{(x_i, y_i)\}_{i=1}^{N_t^*}, \quad (5)$$

where each instance x_i in D_t^{r*} has the s_i^l higher than T_c and N_t^* is the retained instance number.

D. Self-Paced Noisy Knowledge Evolution

After each round of data purification along the training process, LGCP continually introduces more samples as the model converges, some of which are hard ones and disturb the clean knowledge learning in earlier rounds. Besides, the old model inevitably contains erroneous knowledge as shown in Fig. 1(d), resulting in erroneous knowledge accumulation and disturbing the learning of correct knowledge. To address these problems, we introduce a Self-Paced Robust Model Learning scheme and an Erroneous Knowledge Filtering mechanism as follows.

1) *Self-Paced Robust Model Learning*: To enhance the learning stability and inspired by the learning principle of humans, we introduce a self-paced learning paradigm that gradually introduces hard samples within each round. Specifically, we first

rank the samples in D_t^{r*} from high label confidence score s_i^l to low. For the e -th epoch within a round, the proportion of data used for training is

$$r_e = r_0 + e * (1 - r_0) / E, \quad (6)$$

where r_0 is the base data rate, which is set to 0.7 by default. E indicates the total epoch number within each round. The selected training data is denoted as S_e^r . With the above mechanism, the model consolidates the knowledge learned from previous rounds first and gradually extends to more challenging samples, which stabilizes training by reducing parameter perturbations.

2) *Erroneous Knowledge Filtering*: In this section, we aim to filter the features of samples that reflect the learned correct knowledge of the old model M_{t-1} and discard the features that contain erroneous knowledge contained in M_{t-1} . Specifically, given the clean subset D_t^{e*} generated by LGCP, we utilize M_{t-1} to process the D_t^{e*} and the generated features $\{f_{t-1}^i\}_{i=1}^{N_t}$ are fed into the CLS (introduced in Section III-C1), where the obtained scores of each instance is named feature confidence s_i^o . Then, a knowledge distillation weight w_i^o is assigned to each instance x_i by

$$w_i^o = \begin{cases} 0 & s_i^o \leq T_o \\ 1 & s_i^o > T_o \end{cases}, \quad (7)$$

where T_o is the hyperparameter serving as the threshold of s_i^o . Notably, $w_i^o = 0$ indicates that the old knowledge can not correctly process the instance x_i , thus f_{t-1}^i primarily contains erroneous old knowledge and should be discarded during knowledge distillation.

Then we proposed a selective knowledge distillation loss $\mathcal{L}_{sKD}$ to ensure correct old knowledge transfer and active erroneous old knowledge forgetting. Specifically, given a batch of instances $\mathcal{B}$, a maximum subset $\mathcal{B}_o$ where each instance with $w_i^o = 1$ is selected to calculate the inter-instance relation loss as follows:

$$\mathcal{L}_{sKD} = \mathcal{L}_{KD}(\theta_r(M_{t-1}(\mathcal{B}_o), M_t(\mathcal{B}_o))), \quad (8)$$

where θ_r is a relation evaluation function [1] and $\mathcal{L}_{KD}$ is a relation knowledge distillation loss [11].

E. Iterative Label Rectification

Although the above LGCP and SNKE modules could guide the model to learn the correct knowledge, the wrongly labeled data, which can contain abundant information, are discarded. To address this drawback, we propose to rectify the annotated data iteratively along the model learning epochs, ensuring the correct knowledge learning and enhancing informative data utilization simultaneously. Specifically, the label rectification is accomplished by

$$y_t^* = \operatorname{argmax}\{\mathbf{y} * w_l + \hat{\mathbf{y}}_t^r * (1 - w_l)\}, \quad (9)$$

where $\mathbf{y} \in \mathbb{R}^{N_p}$ is a one-hot annotated label embedding and N_p is the annotated person identity number. The label rectification weight w_l aims to fuse the information of annotation and prediction. $\hat{\mathbf{y}}_t^r \in \mathbb{R}^{N_p}$ is the identity prediction vector generated by M_t^r . y_t^* is utilized to replace y in D_t^{r-1} to obtain new dataset D_t^r .

Model Training: During training, our framework can be integrated with existing LReID and LNL methods by introducing their noisy data learning loss $\mathcal{L}_{ReID}$ and our weighted knowledge distillation loss $\mathcal{L}_{sKD}$. Therefore, the overall loss is calculated by:

$$\mathcal{L} = \mathcal{L}_{ReID} + \mathcal{L}_{sKD}. \quad (10)$$

Model Inference: We follow existing methods [1], [9], [11] to use the feature generated by the final model M_T for person matching.

IV. THEORETICAL ANALYSIS AND JUSTIFICATIONS

A. Rationality of Data Confidence Score Generation

In our Cluster-based Label Scoring (CLS) strategy, given N_t image-label pairs $\{(x_i, y_i)\}_{i=1}^{N_t}$, the features are extracted by the model, and the DBSCAN clustering algorithm is adopted to obtain clusters, where the outliers are also gathered as an extra cluster.

Then, according to the clustering results, each image x_i is assigned a cluster-aware label $\tilde{y}_i \in \{1, 2, \dots, N_c\}$ where N_c is the cluster number.

Subsequently, we generate one-hot embedding $\mathbf{l}_i \in \mathbb{R}^{N_c}$ for each x_i according to $\tilde{y}_i$. Therefore, $\mathbf{l}_i$ satisfies:

$$\sum_{j=1}^{N_c} \mathbf{l}_i^{(j)} = 1, \quad (11)$$

where $\mathbf{l}_i^{(j)}$ is the j -th element of $\mathbf{l}_i$, and for arbitrary $j \in \{1, 2, \dots, N_c\}$, $0 \leq \mathbf{l}_i^{(j)} \leq 1$.

Besides, according to (1), we obtain the annotation-aware average cluster label

$$\bar{\mathbf{l}}_i = \frac{1}{n_t^i} \sum_{j=1}^{N_t} \delta(y_i, y_j) \mathbf{l}_j, \quad (12)$$

where $0 \leq \bar{\mathbf{l}}_i^{(k)}$ ($k \in \{1, 2, \dots, N_c\}$) and $\delta(y_i, y_j)$ is a sign function that outputs 1 and 0 when $y_i = y_j$ and $y_i \neq y_j$ respectively. n_t^i indicates the number of images that are annotated with the same label as x_i .

Specifically, n_t^i is obtained by

$$n_t^i = \sum_{j=1}^{N_t} \delta(y_i, y_j), \quad (13)$$

where $n_t^i \geq 1$ since $\delta(y_i, y_j) \geq 0$ and $\delta(y_i, y_j) = 1$ when $i = j$.

Then, the label distance d_i calculation process in (2) of the main paper could be derived as follows:

$$\begin{aligned} d_i &= \|\mathbf{l}_i - \bar{\mathbf{l}}_i\|_2^2 \\ &= \left\| \mathbf{l}_i - \frac{1}{n_t^i} \sum_{j=1}^{N_t} \delta(y_i, y_j) \mathbf{l}_j \right\|_2^2 \\ &= \left\| \frac{n_t^i - 1}{n_t^i} \mathbf{l}_i - \frac{1}{n_t^i} \sum_{j \neq i}^{N_t} \delta(y_i, y_j) \mathbf{l}_j \right\|_2^2 \end{aligned}$$

$$\begin{aligned}
&= \sum_{k=1}^{N_c} \left(\frac{n_t^i - 1}{n_t^i} \mathbf{l}_i^{(k)} - \frac{1}{n_t^i} \sum_{j \neq i}^{N_t} \delta(y_i, y_j) \mathbf{l}_j^{(k)} \right)^2 \\
&\leq \sum_{k=1}^{N_c} \left(\frac{n_t^i - 1}{n_t^i} \mathbf{l}_i^{(k)} \right)^2 + \sum_{k=1}^{N_c} \left(\frac{1}{n_t^i} \sum_{j \neq i}^{N_t} \delta(y_i, y_j) \mathbf{l}_j^{(k)} \right)^2 \\
&= \left(\frac{n_t^i - 1}{n_t^i} \right)^2 \sum_{k=1}^{N_c} (\mathbf{l}_i^{(k)})^2 + \left(\frac{1}{n_t^i} \right)^2 \sum_{k=1}^{N_c} \left(\sum_{j \neq i}^{N_t} \delta(y_i, y_j) \mathbf{l}_j^{(k)} \right)^2 \\
&\leq \left(\frac{n_t^i - 1}{n_t^i} \right)^2 \left(\sum_{k=1}^{N_c} \mathbf{l}_i^{(k)} \right)^2 + \left(\frac{1}{n_t^i} \right)^2 \left(\sum_{j \neq i}^{N_t} \sum_{k=1}^{N_c} \delta(y_i, y_j) \mathbf{l}_j^{(k)} \right)^2 \\
&= \left(\frac{n_t^i - 1}{n_t^i} \right)^2 + \left(\frac{1}{n_t^i} \right)^2 \left(\sum_{j \neq i}^{N_t} \delta(y_i, y_j) \right)^2 \\
&= \left(\frac{n_t^i - 1}{n_t^i} \right)^2 + \left(\frac{1}{n_t^i} \right)^2 (n_t^i - 1)^2 \\
&= 2 \left(\frac{n_t^i - 1}{n_t^i} \right)^2. \tag{14}
\end{aligned}$$

Therefore, $d_i \in [0, 2)$ and simply $s_i = (2 - d_i)/2$ could convert the label distance d_i into label confidence $s_i \in [0, 1]$ (the d_i of the outliers during clustering are set to 2, thereby allowing the possibility of $s_i = 0$).

B. Bounded Error Analysis Across Training Rounds

In this section, we analyze the error propagation behavior of the proposed Continual Self-Paced Dual-Knowledge Purification framework and show that the accumulated error remains bounded across training rounds.

We define the round-level model error at training round r as

$$\varepsilon_r = \mathbb{P}_{x \sim D_t} (\hat{y}_t^r(x) \neq y^*(x)), \tag{15}$$

where $y^*(x)$ denotes the ground-truth identity label and $\hat{y}_t^r(x)$ is the prediction of ILR module. Here, $x \sim D_t$ denotes that x is sampled from the data distribution defined by the dataset D_t .

Theorem 1 (Bounded Error Propagation): Assume that: i) the Local-Global Collaborative Purification (LGCP) module ensures that the probability p_r of selecting an incorrectly labeled sample into the purified set satisfies $p_r \leq c_1 \varepsilon_r$ with $0 < c_1 < 1$; ii) the Self-Paced Noisy Knowledge Evolution (SNKE) strategy introduces samples with ratio $\eta_r \in (0, 1]$ at round r ; and iii) the Erroneous Knowledge Filtering (EKF) mechanism limits the probability of distilling erroneous historical knowledge to at most $\delta \ll 1$. Then, the model error across rounds satisfies

$$\varepsilon_{r+1} \leq \alpha \varepsilon_r + \beta, \quad \text{with } \alpha < 1, \tag{16}$$

where $\alpha = 1 - \eta_r(1 - c_1) + \delta$, and β denotes an irreducible error term induced by intrinsic label noise. Consequently,

$$\limsup_{r \rightarrow \infty} \varepsilon_r \leq \frac{\beta}{1 - \alpha}. \tag{17}$$

Proof Sketch: At round $r + 1$, the error originates from residual noisy supervision and propagated historical knowledge. LGCP bounds the noise level of the purified training set, while SNKE limits the influence of uncertain samples. Meanwhile, EKF suppresses erroneous historical knowledge during distillation. Combining these effects yields the contractive recursion in (16), which guarantees bounded error accumulation. The detailed derivation of the conditions and conclusions of Theorem 1 is provided in the supplementary material. $\square$

V. EXPERIMENTS

A. Benchmark

In this paper, following the existing LReID [9] and Noisy ReID [14] works, we establish a new Noisy LReID benchmark (NLReID) as described below.

Datasets: NLReID contains 12 ReID datasets. Among them, 5 datasets are used for lifelong training (Market1501 [74], DukeMTMC-reID [75], CUHK-SYSU [76], MSMT17-V2 [77], and CUHK03 [78]), and the other 7 test datasets are used evaluate model's generalizability (CUHK01 [79], CUHK02 [80], VIPeR [81], PRID [82], i-LIDS [83], GRID [84], and SenseReID [85]). The identity configurations of all datasets follow the settings established in LReID [9]. Detailed dataset statistics, including the numbers of identities and images in the training, gallery, and query splits, are visualized in Fig. 4.

Label noise generation: We follow the settings in [14] and consider two types of noisy annotations:

1) *Random noise.* A certain percentage (10%, 20%, 30%) of training samples are randomly reassigned to other identities. This setting simulates label corruption caused by arbitrary annotation errors.

2) *Patterned noise.* A certain proportion (10%, 20%, 30%) of training images are mislabeled by assigning them the identity of their most similar instances from other identities. For each dataset (i.e., Market-1501, CUHK-SYSU, and DukeMTMC), an individual ReID model is trained using all clean labels. Then, this model was employed to measure sample similarity. This setting is designed to simulate realistic annotation errors that occur when instances from different identities exhibit highly similar semantic patterns.

Evaluation Metrics: Following existing LReID works [9], [11], [36], we adopt mean Average Precision (mAP) and Rank-1 accuracy (R@1) to evaluate model performance on each seen and unseen dataset. To further assess both lifelong learning ability and cross-domain generalization, we also report the average mAP and R@1 across all seen domains and across all unseen domains, respectively. These are denoted as Seen-Avg mAP, Seen-Avg R@1, Unseen-Avg mAP, and Unseen-Avg R@1.

B. Implementation Details

Our method is built based on LSTKC [1]. During training, the first dataset is trained for 80 epochs, and the subsequent datasets are trained for 60 epochs. 32 identities with 4 images for each identity are sampled as a mini-batch. The learning rate and weight decay are set as 0.008 and 0.0001, respectively, and an SGD optimizer is adopted. The hyperparameters T_c, T_o ,

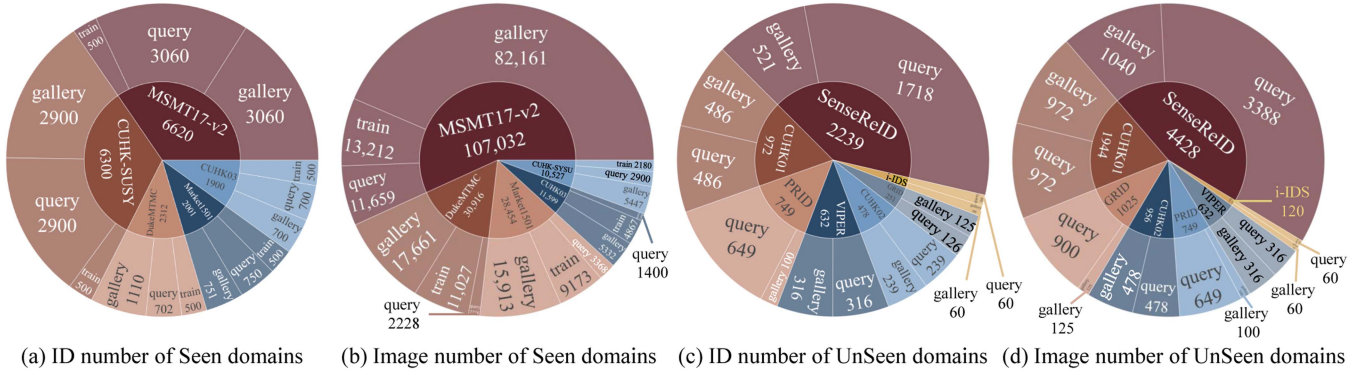


Fig. 4. The statistics of identity (ID) and image numbers for seen and unseen domains in the NLReID benchmark. The varying scales across domains follow the LReID benchmark convention [9] to mimic dynamic data volumes in practical applications. Noisy-label ratios (10%, 20%, 30%) are described in Section V and evaluated in Tables II and III.

w_l , and w_c are set to 0.5, 0.3, 0.3, and 0.5, respectively. For the DBSCAN algorithm, the parameter ϵ [68] and minimum samples are set to 0.5 and 3, respectively. All the experiments are conducted on a single NVIDIA GPU using the PyTorch framework. Unless otherwise stated, results are reported under the benchmark setting with 30% random label noise.

C. The Comparison Methods

To comprehensively evaluate our method, extensive state-of-the-art LReID approaches (PatchKD [11], KRKC [73], LSTKC [1], DKP [12], DASK [37] and LSTKC++ [38]) are compared. Besides, the recent continual learning (CL) methods LwF [69], LKD [70], LEAD [71], and PASS++ [72] are also included. In addition, we combine the efficient and state-of-the-art LReID method, LSTKC, with the latest noisy label learning methods (LCNL [17], DICS [53], CORE [14], PNET [51] and SELF [14]) to adapt them to the Noisy LReID scenario. For fairness, all methods are implemented using their official codebases, and a consistent ResNet-50 backbone, data noise ratio, and noise type are adopted across all comparisons.

D. Comparison With State-of-The-Art Methods

Tables II and III visualize the results on the NLReID benchmark under random and patterned noise, respectively. The best results in each setting are highlighted in **RED**, and the second-best in **BLUE**.

Compared to CL Methods: As shown in Tables II and III, continual learning (CL) approaches exhibit inferior performance compared to state-of-the-art LReID, LNL, and Noisy LReID methods. This performance gap arises because most CL methods are originally designed for image classification with clean labels. In Noisy LReID scenarios, the supervision conditions are naturally noisy and fine-grained. As a result, their ability to gain new knowledge and remember old knowledge is greatly limited.

Compared to LReID Methods: As shown in Tables II and III, our CKP and CSDP both achieve substantially better average performance on both seen and unseen domains compared to existing LReID methods. Specifically, when the random noise

ratio increases from 10% to 30%, state-of-the-art LReID methods suffer severe performance drops, with average mAP and R@1 degradation of **12%-24%**. In contrast, CKP and CSDP experience at most **6.1%** and **2.6%** degradation, respectively. The robustness of CKP results from its local relation-based data purification and erroneous knowledge filtering designs, while CSDP further improves performance by incorporating global classification clues that improve clean data utilization and a self-paced learning mechanism that stabilizes model training.

Compared to LNL Methods: Tables II and III further show that CKP and CSDP consistently outperform state-of-the-art LNL approaches (CORE, DICS, SELF, PNET). Under 30% random noise, CKP improves average mAP/R@1 by **12.9%/14.2%** on seen domains and **10.0%/8.9%** on unseen domains. Under 30% patterned noise, CKP yields improvements of **9.0%/9.1%** and **6.1%/5.7%**, respectively. Our enhanced CSDP achieves even larger gains: **17.6%/19.6%** and **13.3%/15.4%** under 30% random noise, and **12.8%/13.1%** and **9.4%/8.9%** under 30% patterned noise. These results highlight that LNL methods, while effective in static settings, fail to address catastrophic forgetting in continual learning, whereas our framework explicitly mitigates both forgetting and erroneous knowledge accumulation.

Compared to Noisy LReID Methods: We incorporate the anti-forgetting strategy from the LReID method LSTKC into LNL methods, resulting in the Noisy LReID variants DICS[†], CORE[†], and LCNL[†]. Among them, CORE[†] achieves the best average performance across different settings. As shown in Table II, compared to CORE[†] under random noise, CKP achieves improvements of **1.3%/0.9%**, **5.0%/5.9%**, and **11.3%/13.3%** in average mAP/R@1 on seen domains under noise ratios of **10%**, **20%**, and **30%**, respectively. Similarly, CSDP achieves improvements of **2.2%/1.8%**, **8.3%/9.6%**, and **16.0%/18.7%** under the same settings. Moreover, CKP and CSDP achieve **3.1%-9.7%** and **3.4%-13.0%** gains, respectively, on unseen domains as the noise ratio increases from 10% to 30%. As shown in Table III, under patterned noise, both CKP and CSDP demonstrate progressively larger performance gains as the noise ratio increases. These improvements can be primarily attributed to their ability to leverage local relational information, which

TABLE II
RESULTS UNDER THE RANDOM NOISE. [†] INDICATES THAT THE STATE-OF-THE-ART LNL METHOD IS COMBINED WITH THE STEE-OF-THE-ART ANTI-FORGETTING STRATEGY OF LREID.

Metric	Type	Method	Market-1501			CUHK-SYSU			DukeMTMC			MSMT17			CUHK03			Seen-Avg			Unseen-Avg		
			10%	20%	30%	10%	20%	30%	10%	20%	30%	10%	20%	30%	10%	20%	30%	10%	20%	30%	10%	20%	30%
mAP	CL	LwF [69]	49.2	29.1	19.2	66.2	53.7	47.7	19.0	9.8	7.5	3.6	1.9	1.3	17.3	8.9	6.1	31.1	20.7	16.4	35.4	25.6	20.2
		LKD [70]	48.5	35.2	24.8	77.3	72.5	68.2	32.4	19.8	14.5	7.8	5.2	3.9	14.6	7.2	4.7	36.1	28.0	23.2	44.6	36.2	32.2
		LEAD [71]	41.5	28.7	17.5	77.7	71.4	66.4	32.6	17.0	10.9	9.2	5.0	3.2	15.8	6.7	4.2	35.4	25.7	20.4	45.4	35.2	30.1
		PASS++ [72]	52.9	36.3	23.7	65.1	58.2	54.0	12.3	7.0	5.1	2.6	1.8	1.3	3.9	1.5	1.1	27.4	21.0	17.0	28.7	22.3	19.9
	LReID	PatchKD [11]	51.0	30.6	23.9	66.8	58.0	52.6	19.3	11.9	8.4	3.6	2.1	1.5	18.1	10.3	6.4	31.8	22.6	18.6	35.0	27.3	23.8
		KRKC [73]	29.4	20.1	14.8	71.7	65.3	59.2	23.3	15.3	10.3	5.1	3.1	2.2	35.3	20.9	12.7	33.0	24.9	19.8	42.5	35.4	29.0
		DKP [12]	43.2	28.1	20.4	78.4	71.0	66.7	34.8	22.5	15.4	12.8	7.7	5.5	21.2	10.7	6.6	38.1	28.0	22.9	47.4	37.6	32.5
		LSTKC [1]	41.4	35.1	25.8	78.7	74.2	67.4	39.7	19.4	9.3	14.8	5.2	2.6	27.8	15.3	6.9	40.5	29.8	22.4	47.2	37.0	30.0
		DASK [1]	42.1	25.9	19.9	71.3	63.2	63.6	43.5	26.0	20.4	16.7	9.3	7.6	31.1	12.9	6.9	40.9	27.5	23.7	53.7	38.9	36.0
		LSTKC++ [38]	50.4	26.6	19.2	81.6	73.8	69.2	42.5	19.0	12.9	11.3	6.4	3.8	30.0	2.1	1.4	43.2	25.6	21.3	52.4	34.6	28.4
	LNL	CORE [14]	35.5	29.4	23.7	75.7	73.7	69.2	36.3	30.4	22.1	12.5	9.9	6.5	41.8	35.2	24.9	40.4	35.7	29.3	51.1	47.3	40.4
		DICS [53]	34.0	23.1	15.7	72.1	70.1	63.6	34.8	22.3	13.1	12.2	7.3	4.5	35.5	18.5	9.3	37.7	28.3	21.2	48.0	38.8	32.2
		SELF [14]	32.3	24.0	15.6	71.5	69.1	62.8	34.1	21.2	11.8	10.6	7.2	3.8	44.4	30.0	16.2	38.6	30.3	22.1	49.8	41.2	31.8
		PNET [51]	23.2	19.0	12.7	69.2	65.5	58.5	21.0	14.6	8.6	6.8	4.7	2.5	38.2	29.4	19.8	31.7	26.6	20.4	43.0	36.4	30.0
	Noisy LReID	DICS [†] [53]	38.3	38.3	29.4	73.4	73.4	67.4	37.1	37.1	14.0	8.8	8.8	3.2	13.1	13.1	6.4	34.1	34.1	24.1	41.5	41.5	32.7
		CORE [†] [14]	48.9	45.2	37.8	81.6	80.2	74.7	46.2	37.1	21.3	18.2	10.5	5.5	35.5	23.9	15.3	46.1	39.4	30.9	52.9	47.2	40.7
		LCNL [†] [17]	28.9	23.5	16.2	69.2	67.9	63.1	29.4	20.0	11.6	9.7	6.8	4.3	38.1	27.8	18.6	35.1	29.2	22.8	45.1	38.4	33.8
		PNET [†] [51]	46.8	30.7	23.5	78.6	69.7	66.2	18.3	6.9	5.6	5.4	2.1	1.6	12.4	6.4	5.0	32.3	23.2	20.4	40.3	30.8	26.3
		SELF [†] [14]	46.7	41.7	33.3	79.8	77.7	71.3	44.4	29.9	12.8	17.7	8.2	3.6	34.5	20.2	10.3	44.7	35.5	26.2	52.9	43.5	34.4
		CKP (Ours)	48.7	44.5	42.2	80.8	80.3	78.6	47.3	44.4	42.1	18.1	16.4	14.6	42.0	36.3	33.5	47.4	44.4	42.2	56.0	51.4	50.4
		CSDP (Ours)	51.6	50.2	48.1	80.7	80.9	80.0	47.6	46.6	46.2	18.0	18.1	18.2	43.5	42.8	42.0	48.3	47.7	46.9	56.3	55.4	53.7
R@1	CL	LwF [69]	74.2	55.9	43.0	69.9	57.4	51.4	35.5	20.3	17.9	11.3	6.7	5.0	17.4	8.6	5.5	41.7	29.8	24.6	28.9	20.5	15.5
		LKD [70]	71.5	59.1	47.1	80.3	75.8	71.6	48.5	32.1	25.8	21.2	15.1	11.8	12.6	5.4	3.6	46.8	37.5	32.0	38.2	30.5	27.6
		LEAD [71]	65.2	52.7	38.0	81.2	75.0	70.1	49.9	29.5	21.5	25.4	15.3	10.5	13.6	5.3	3.5	47.1	35.5	28.7	39.5	29.6	25.6
		PASS++ [72]	75.5	62.7	49.0	69.1	62.0	57.4	23.7	17.1	13.1	9.6	6.8	5.3	3.8	1.0	0.7	36.3	29.9	25.1	23.4	16.6	14.8
	LReID	PatchKD [11]	74.2	56.9	49.5	70.5	61.7	56.0	34.1	23.1	18.1	11.0	7.2	5.5	17.2	9.8	5.9	41.4	31.7	27.0	29.6	21.7	18.9
		KRKC [73]	54.0	42.0	34.4	75.1	68.8	63.4	38.4	28.0	19.8	14.9	9.8	7.5	37.1	19.6	11.1	43.9	33.6	27.2	36.9	29.4	24.3
		DKP [12]	68.7	54.3	45.7	81.1	74.8	70.4	53.9	39.5	30.4	32.5	23.3	18.2	20.4	9.9	5.5	51.3	40.4	34.0	41.0	32.0	26.8
		LSTKC [1]	66.8	60.5	50.2	81.3	77.6	70.9	59.3	34.0	19.5	35.5	16.2	8.8	27.7	14.6	6.2	54.1	40.6	31.1	40.3	30.7	25.0
		DASK [37]	67.9	55.0	46.8	73.8	67.6	68.4	65.1	48.5	40.0	42.3	31.7	27.3	32.3	13.6	6.1	56.3	43.3	37.7	47.0	32.8	30.6
		LSTKC++ [38]	73.5	52.4	42.9	83.9	77.2	73.2	61.0	33.3	24.3	29.3	19.6	13.1	30.5	1.5	0.9	55.6	36.8	30.9	45.5	29.4	22.8
	LNL	CORE [14]	63.0	55.8	49.8	78.7	77.0	72.6	56.7	48.9	38.8	32.9	26.9	19.0	43.1	36.2	24.1	54.9	49.0	40.9	44.7	41.7	34.5
		DICS [53]	58.6	48.5	38.2	74.9	74.2	67.9	55.7	40.6	27.2	33.1	23.0	16.6	36.3	18.3	8.6	51.7	40.9	31.7	42.0	33.4	27.3
		SELF [14]	58.7	49.6	36.8	74.7	73.2	66.5	53.3	37.0	25.1	29.5	21.7	13.5	47.9	30.1	15.9	52.8	42.3	31.6	43.7	35.8	26.5
		PNET [51]	47.2	41.3	30.1	72.9	69.1	62.3	36.5	27.3	18.8	19.9	14.6	8.8	40.9	30.8	19.4	43.5	36.6	27.9	37.1	30.7	24.7
	Noisy LReID	DICS [†] [53]	65.6	65.6	56.4	76.8	76.8	70.9	57.3	57.3	28.1	26.7	26.7	12.4	12.8	12.8	6.2	47.8	47.8	34.8	34.9	34.9	27.3
		CORE [†] [14]	72.6	69.9	62.8	83.9	83.0	77.4	63.8	55.5	37.5	41.0	27.1	16.8	36.4	22.9	14.7	59.5	51.7	41.8	46.0	40.9	35.3
		LCNL [†] [17]	55.3	48.8	37.6	72.7	71.8	66.9	49.1	36.3	23.1	27.6	21.0	14.7	39.7	29.4	18.9	48.9	41.5	32.2	38.4	32.6	28.5
		PNET [†] [51]	70.4	55.8	46.5	81.4	73.0	69.6	31.7	16.2	13.6	15.3	6.9	5.4	12.3	5.4	4.4	42.2	31.5	27.9	33.9	25.1	20.5
		SELF [†] [14]	71.7	66.6	59.0	82.4	80.4	74.9	63.5	48.3	26.3	40.6	22.6	12.1	35.6	19.7	9.0	58.8	47.5	36.2	46.0	37.4	29.1
		CKP (Ours)	71.8	68.1	66.9	83.2	83.0	81.0	64.7	62.1	58.9	40.1	37.5	34.7	42.4	37.1	34.1	60.4	57.6	55.1	49.5	44.6	43.4
		CSDP (Ours)	73.2	72.5	71.6	82.9	83.0	81.9	65.7	65.8	64.0	40.4	40.6	40.9	44.1	44.6	43.9	61.3	61.3	60.5	49.5	47.8	46.9

effectively discriminates between noisy and clean labels. As a result, correct knowledge is more accurately mined, while the retention of erroneous knowledge is mitigated in highly noisy scenarios.

Statistical comparison: To further verify the statistical significance of our method, we evaluate CKP and CSDP with the state-of-the-art methods DASK and CORE[†] over three independent runs, and report the mean and standard deviation in Fig. 5. The results show that CKP and CSDP consistently outperform the existing methods. In particular, CSDP achieves improvements of **1.0%–2.8%**, **5.1%–10.0%**, and **6.7%–18.1%** under noise ratios of 10%, 20%, and 30%, respectively. These results verify the robustness of our proposed continual self-paced dual-knowledge purification mechanism.

E. Ablation Studies

In this section, we evaluate and discuss the effectiveness of the proposed model components. All the experiments are conducted under 30% random noise setting by default.

Ablations on Model Components: In Table IV, we provide the ablation studies on the key components of our CSDP model. Specifically, we start with a CORE[†] baseline and progressively integrate the proposed LGCP, SNKE, and ILR modules. Specifically, LGCP aims to filter noisy samples, SNKE focuses on suppressing noisy historical features, and ILR is designed to rectify erroneous labels. The results show that each module independently contributes to performance improvement, while their joint integration further boosts overall results. Notably, LGCP

TABLE III
RESULTS UNDER THE PATTERNED NOISE. [†] INDICATES THAT THE STATE-OF-THE-ART LNL METHOD IS COMBINED WITH THE STEE-OF-THE-ART ANTI-FORGETTING STRATEGY OF LReID.

Metric	Method	Market-1501			CUHK-SYSU			DukeMTMC			MSMT17			CUHK03			Seen-Avg			Unseen-Avg			
		10%	20%	30%	10%	20%	30%	10%	20%	30%	10%	20%	30%	10%	20%	30%	10%	20%	30%	10%	20%	30%	
mAP	CL	LwF [69]	47.8	33.2	17.4	60.8	47.3	32.5	22.6	11.6	5.9	4.4	2.0	1.0	11.9	5.7	3.3	29.5	20.0	12.0	35.4	23.8	19.0
		LKD [70]	53.0	44.1	36.3	78.1	76.0	73.1	35.6	26.5	21.1	9.1	6.6	5.2	16.6	10.2	8.6	38.5	32.7	28.9	47.2	41.4	39.0
		LEAD [71]	44.3	34.8	27.3	75.5	71.5	67.1	33.9	22.4	15.5	10.4	6.1	4.3	17.4	10.4	6.9	36.3	29.0	24.2	46.8	39.0	35.6
		PASS++ [72]	62.4	51.7	43.3	68.5	64.5	60.9	15.4	11.3	8.4	3.5	2.5	1.9	5.3	3.8	2.6	31.0	26.8	23.4	33.5	28.2	26.2
	LReID	PatchKD [11]	50.6	34.2	18.9	62.1	48.2	34.7	22.1	12.1	5.9	4.5	2.2	1.1	11.4	5.5	2.9	30.1	20.4	12.7	35.1	25.5	18.5
		KRKC [73]	31.9	22.3	18.0	73.1	66.7	63.5	27.1	16.2	10.4	5.7	3.4	2.6	39.1	24.8	16.1	35.4	26.7	22.1	45.4	36.9	31.9
		DKP [12]	46.1	33.7	27.4	80.3	73.8	70.1	37.4	26.1	20.3	14.2	9.2	7.2	24.8	14.9	9.3	40.6	31.5	26.9	50.5	41.2	36.7
		LSTKC [1]	40.3	36.0	33.2	78.3	75.4	71.6	38.3	30.6	18.0	15.2	8.0	4.4	30.2	20.0	13.2	40.5	34.0	28.1	50.8	42.2	37.9
		DASK [37]	47.5	34.6	23.8	73.1	64.8	67.2	44.5	31.9	24.4	18.1	10.8	8.4	31.8	23.2	8.8	43.0	33.0	26.5	55.0	45.2	39.6
		LSTKC++ [38]	57.1	48.9	36.5	80.9	75.7	73.3	44.3	35.4	24.3	12.6	6.7	4.6	34.9	11.1	2.8	46.0	35.6	28.3	57.0	42.4	35.6
	LNL	CORE [14]	35.7	33.4	30.1	75.8	74.8	73.3	36.9	30.9	29.2	13.5	11.1	9.3	42.4	37.7	31.5	40.9	37.6	34.7	52.2	46.5	45.0
		DICS [53]	32.3	26.1	19.1	69.3	71.3	67.3	34.2	26.6	16.1	10.8	9.5	5.7	35.8	23.6	13.2	36.5	31.4	24.3	45.9	43.7	35.8
		SELF [14]	33.5	29.6	23.6	71.3	71.2	67.8	33.2	28.5	18.4	11.4	9.3	6.3	44.0	35.4	24.7	38.7	34.8	28.2	48.7	44.0	38.3
		PNET [51]	23.9	19.2	18.3	69.4	66.6	63.3	19.2	16.1	12.4	6.5	5.0	3.6	37.8	31.4	24.6	31.4	27.7	24.4	43.1	38.7	34.7
	Noisy LReID	DICS [†] [53]	40.5	39.5	31.4	77.4	75.0	68.1	40.1	39.5	33.8	16.4	9.0	6.9	33.7	21.1	16.4	41.6	36.8	31.3	52.7	47.4	41.1
		CORE [†] [14]	49.0	49.6	45.4	80.8	80.0	79.0	46.4	42.2	38.7	19.2	9.2	10.8	39.5	32.2	21.7	47.0	42.6	39.1	55.8	48.6	47.6
		LCNL [†] [17]	29.6	25.9	21.1	68.4	67.6	65.6	29.2	23.0	19.6	9.7	7.6	6.2	38.5	31.5	23.6	35.1	31.1	27.2	45.1	40.7	38.0
		PNET [†] [51]	49.3	35.8	28.5	80.0	73.2	68.2	24.6	9.8	6.6	6.6	2.5	1.7	12.3	7.6	8.5	34.6	25.8	22.7	43.4	32.5	30.8
		SELF [†] [14]	46.0	41.4	42.0	79.1	78.3	76.6	45.0	40.1	33.9	18.0	13.8	8.4	38.7	27.0	17.5	45.4	40.1	35.7	52.8	47.1	44.4
		CKP (Ours)	50.1	46.9	44.1	81.0	79.9	78.9	47.2	45.2	43.1	18.3	17.1	15.7	41.9	39.9	36.6	47.7	45.8	43.7	57.3	54.4	51.1
		CSDP (Ours)	51.5	50.0	50.9	81.4	80.4	80.4	47.3	46.5	46.3	18.2	18.0	17.9	44.0	43.1	42.1	48.5	47.6	47.5	56.5	56.1	54.4
R@1	CL	LwF [69]	72.0	59.8	37.5	63.8	49.3	32.7	42.1	25.0	14.5	14.5	7.0	4.4	12.8	6.7	3.8	41.0	29.6	18.6	28.3	17.8	13.7
		LKD [70]	74.8	66.1	58.7	80.6	78.8	76.6	52.4	40.6	34.6	24.0	18.3	14.6	15.3	8.7	7.1	49.4	42.5	38.3	40.2	35.0	32.5
		LEAD [71]	68.3	58.2	50.0	78.8	75.6	71.3	52.8	36.9	29.2	28.3	18.5	13.5	16.6	9.9	5.7	49.0	39.8	33.9	40.7	33.4	29.8
		PASS++ [72]	81.5	74.1	67.8	72.1	67.6	64.7	28.7	23.3	17.5	11.8	8.7	6.9	4.5	3.1	1.9	39.7	35.4	31.8	26.7	22.1	20.6
	LReID	PatchKD [11]	74.6	60.2	40.8	64.7	50.9	36.4	40.8	26.8	15.7	14.2	7.7	4.6	11.6	6.1	3.2	41.2	30.3	20.1	28.9	20.1	13.1
		KRKC [73]	57.5	44.7	38.5	76.5	70.4	67.0	43.4	28.6	20.0	16.2	10.9	8.6	40.3	23.9	15.9	46.8	35.7	30.0	38.5	32.2	27.0
		DKP [12]	71.1	60.2	53.8	82.9	77.2	74.1	55.4	42.8	36.8	34.8	25.9	21.5	24.8	14.3	7.6	53.8	44.1	38.8	43.4	35.9	30.7
		LSTKC [1]	65.2	59.9	58.1	80.8	78.0	74.8	56.3	48.6	32.4	36.8	22.5	13.6	31.3	20.5	12.9	54.1	45.9	38.4	44.3	36.1	32.2
		DASK [37]	73.4	61.2	51.1	75.7	67.9	71.2	65.4	54.5	45.1	44.3	33.7	28.9	33.6	23.4	7.1	58.5	48.1	40.7	48.5	38.3	33.5
		LSTKC++ [38]	77.5	71.7	61.2	82.6	78.2	77.1	62.8	53.8	40.5	31.7	20.8	15.1	35.8	10.3	2.2	58.1	47.0	39.2	50.5	36.2	30.1
	LNL	CORE [14]	63.4	60.1	57.1	78.7	77.9	76.7	57.3	50.7	46.7	34.1	29.6	25.6	43.8	38.2	32.4	55.5	51.3	47.7	45.9	40.1	38.9
		DICS [53]	57.3	51.9	43.2	72.5	75.0	71.3	55.1	44.5	31.2	29.7	28.4	19.3	36.4	24.1	13.0	50.2	44.8	35.6	38.7	37.3	30.3
		SELF [14]	61.0	56.0	48.1	74.3	74.3	71.9	53.9	47.1	33.7	30.6	27.5	19.5	44.6	37.0	25.7	52.9	48.4	39.8	41.6	38.3	32.9
		PNET [51]	48.7	41.1	40.2	72.9	70.3	67.2	34.9	29.6	24.0	19.7	15.5	12.1	38.9	31.5	25.4	43.0	37.6	33.8	37.8	33.1	29.1
	Noisy LReID	DICS [†] [53]	65.7	64.5	58.9	79.5	78.0	71.9	58.8	59.2	54.8	40.1	26.8	22.9	35.1	21.6	16.4	55.8	50.0	45.0	46.5	40.9	35.0
		CORE [†] [14]	72.6	72.8	69.9	82.8	82.3	81.6	65.0	59.0	56.3	43.4	24.6	28.1	40.5	31.9	21.2	60.9	54.1	51.4	48.7	42.1	40.8
		LCNL [†] [17]	55.8	51.6	47.0	71.5	71.1	69.5	49.2	41.1	35.7	27.2	23.0	19.4	39.6	33.3	24.6	48.7	44.0	39.2	38.4	34.9	32.3
		PNET [†] [51]	72.2	59.8	51.4	82.6	76.5	72.0	40.6	19.8	15.1	18.4	8.3	6.0	11.1	7.2	7.3	45.0	34.3	30.4	36.8	27.1	25.0
		SELF [†] [14]	71.0	66.1	67.0	81.3	81.1	79.9	64.9	59.3	53.5	41.5	34.7	23.9	40.9	26.9	16.9	59.9	53.6	48.2	45.9	40.8	38.5
		CKP (Ours)	73.2	70.8	67.5	82.7	82.3	81.4	65.8	62.5	61.2	41.0	38.9	36.7	42.7	40.8	37.0	61.1	59.1	56.8	50.1	47.5	44.6
		CSDP (Ours)	74.0	73.4	73.7	83.3	82.4	82.8	64.4	64.0	63.9	40.3	40.7	40.4	45.6	44.5	43.4	61.5	61.0	60.8	49.8	49.4	47.8

provides the most significant gain, highlighting the importance of improving data reliability in mitigating the effects of label noise.

We further conduct ablation studies on the conference version CKP, which comprises Cluster-aware Data Purification (CDP), Erroneous Knowledge Filtering (EKF), and Iterative Label Rectification (ILR). The results are shown in Fig. 6. When replacing CDP and EKF with the improved LGCP and SNKE modules, our method shows consistent improvements on both seen and unseen domains. The improvement brought by LGCP stems from its local-global information utilization mechanism, which enhances noisy data recognition. Meanwhile, by leveraging a self-paced learning paradigm, SNKE improves training stability and accelerates correct knowledge

accumulation. Moreover, LGCP and SNKE are complementary, and their joint integration yields further improvements of **4.7%/5.4%** Seen-Avg mAP/R@1 and **3.3%/3.5%** UnSeen-Avg mAP/R@1, forming our CSDP.

Ablations on Hyperparameters: We analyze the effects of hyperparameters T_c , T_o , w_l , and w_c in Fig. 7. Fig. 7(a) shows that a relatively large T_c improves overall performance. This is because a larger T_c results in higher data purity, thus reducing the influence of noisy labels. However, a larger T_c also results in fewer clean labels, thus degrading the acquisition capacity. Overall, $T_c = 0.5$ achieves the best balance. In Fig. 7(b), an optimal T_o is relatively small. This is because some features simultaneously encode correct and erroneous knowledge. A larger T_o helps new knowledge learning but intensifies

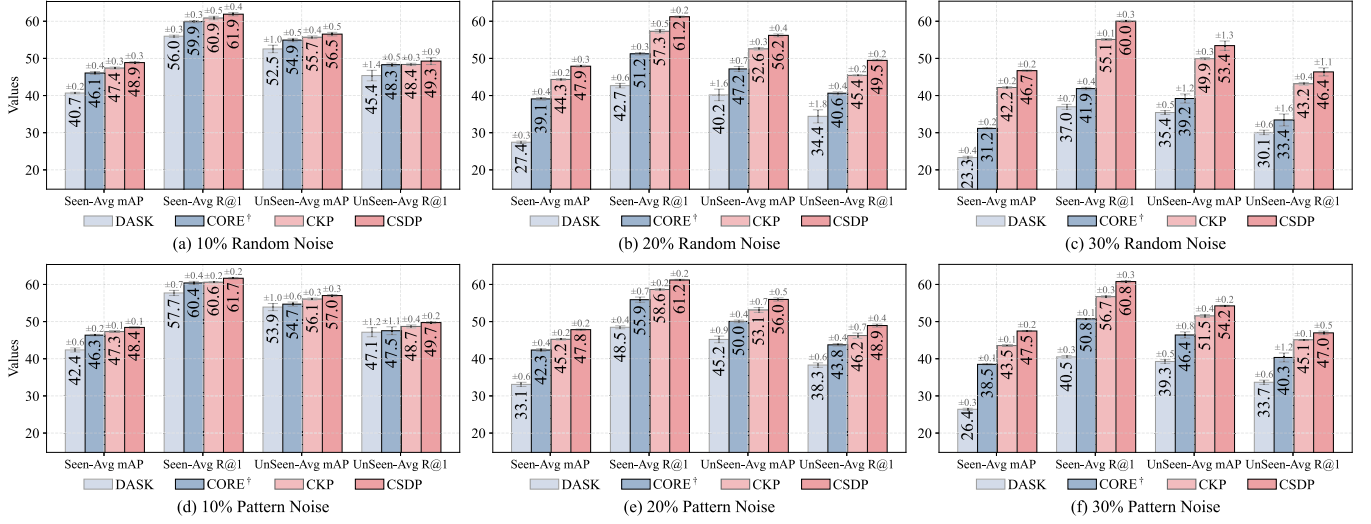


Fig. 5. Statistical comparison of different methods under diverse noise configurations. For each configuration, all methods are evaluated over three independent runs, and the mean and standard deviation are reported.

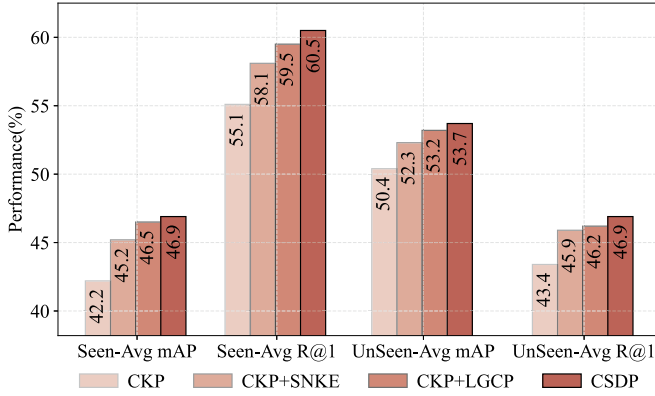


Fig. 6. Module ablation studies of the proposed CSDP. CSDP is equivalent to CKP+LGCP+SNKE. CKP+LGCP and CKP+SNKE denote using the improved version to replace the label purification and knowledge evolution modules, respectively.

TABLE IV
ABLATION STUDY OF DIFFERENT COMPONENTS IN OUR CSDP. LGCP IS THE LOCAL-GLOBAL COLLABORATIVE PURIFICATION MODULE. SNKE IS THE SELF-PACED NOISY KNOWLEDGE EVOLUTION MODULE. ILR IS THE ITERATIVE LABEL RECTIFICATION MODULE.

Baseline	LGCP	SNKE	ILR	Seen-Avg mAP	Seen-Avg R@1	Unseen-Avg mAP	Unseen-Avg R@1
✓				30.9	41.8	40.7	35.3
✓	✓			40.6	52.4	47.9	41.3
✓		✓		37.7	50.5	45.8	39.2
✓			✓	34.5	45.7	43.5	37.4
✓	✓		✓	43.5	54.7	49.2	43.1
✓	✓	✓	✓	46.9	60.5	53.7	46.9

forgetting. $T_o = 0.3$ provides the best balance between new and old knowledge. Fig. 7(c) indicates that a relatively low w_l is optimal for label rectification. This is because model predictions are generally more reliable than annotated labels,

while annotated labels still contain useful cues that complement imperfect predictions. Finally, $w_l = 0.3$ achieves the best balance. In Fig. 7(d), $w_c = 0$ and $w_c = 1$ correspond to using only global or local relations, respectively. The results show that $w_c = 0.5$ achieves the best performance. This is because relying solely on either source is limited, and the complementary semantics between local and global information are essential for effective noisy data purification. Besides, the choice of r_0 has been discussed in our conference version CKP. Specifically, when $r_0 < 5$, the model performance remains relatively stable, as clustering is performed before the model fully converges. In contrast, when $r_0 > 5$, performance gradually degrades because the model begins to overfit noisy labels. Therefore, setting $r_0 = 5$ achieves the best trade-off between stability and robustness.

Ablations on Parameters of DBSCAN: We further conduct a sensitivity analysis of the DBSCAN algorithm's key parameters. As illustrated in Fig. 8, ϵ denotes the maximum distance within which two samples are considered neighbors. The results indicate that model performance remains relatively stable for $\epsilon \in [0.3, 0.7]$. When $\epsilon < 0.3$ or $\epsilon > 0.7$, the performance drops sharply. This is because a larger ϵ can introduce more clustering noise. While a smaller ϵ results in more outliers and reduces the utilization of training samples. Additionally, 'Min sample' represents the minimum number of points required to form a cluster. 'Min sample' exhibits minimal impact on performance, suggesting that the model is robust to variations in this parameter.

Ablations on Clustering Algorithms: As shown in Fig. 9, we conduct ablation studies on the clustering algorithms adopted in the proposed CSDP framework. The evaluated methods include DBSCAN [68], HDBSCAN [87], and Spectral Clustering [88]. The results demonstrate that all three clustering algorithms consistently improve performance, which verifies the robustness of our framework with respect to the choice of clustering strategy. Among them, DBSCAN achieves slightly superior performance compared to HDBSCAN and Spectral Clustering. This advantage can be attributed to DBSCAN's ability to effectively capture

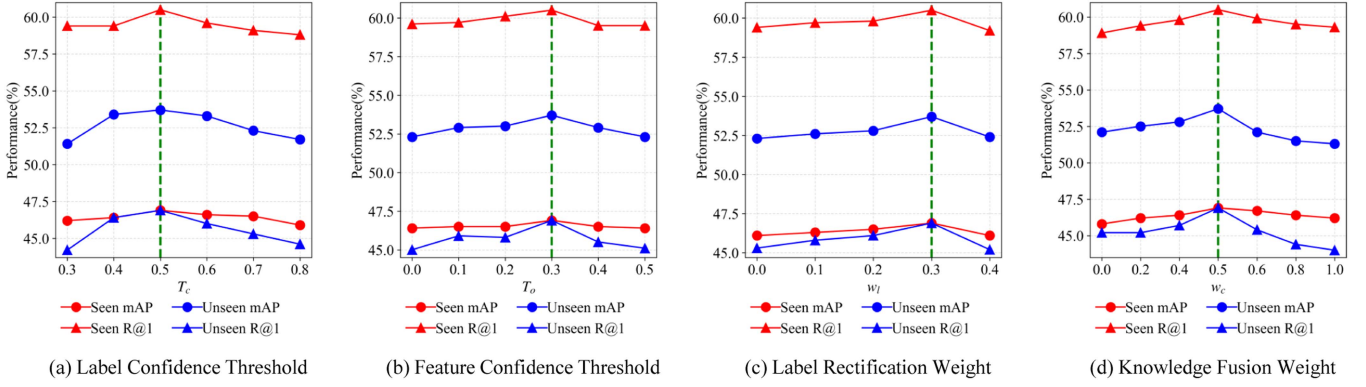


Fig. 7. Ablation studies on the sensitivity of hyperparameters T_c , T_o , w_l , and w_c . The dashed green lines indicate the default values used in our experiments.

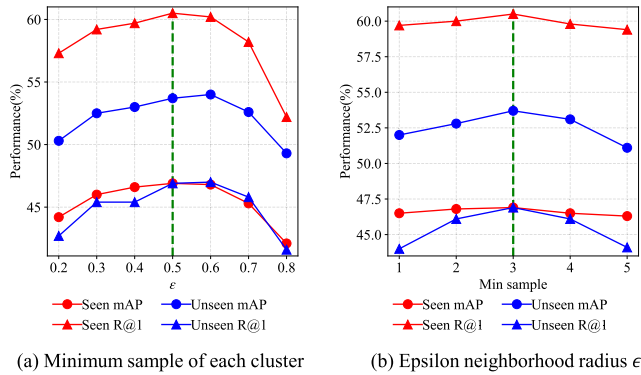


Fig. 8. Ablation studies on the sensitivity of parameters in DBSCAN.

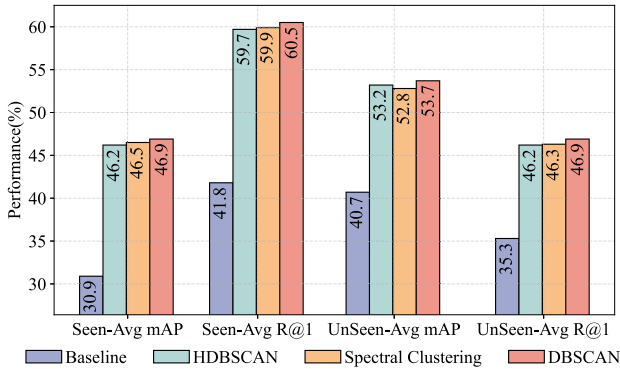


Fig. 9. Ablation studies on different clustering algorithms.

local neighborhood relationships, which are complementary to the global discriminative cues provided by the classification head. In contrast, HDBSCAN and Spectral Clustering place more emphasis on global data structures, which may limit their capacity to exploit fine-grained local relations.

Evaluation of Training Cost: As shown in Table V, we compare the training costs of CKP and CSDP with state-of-the-art methods. Overall, both CKP and CSDP incur **comparable computational overhead** to existing approaches, *e.g.*, DKP, DASK, and CORE[†]. In particular, compared to our default baseline CORE[†], CSDP requires an additional 1.3 hours of training time

TABLE V
COMPUTING EFFICIENCY COMPARISON WITH EXISTING LREID METHODS

Method	Train Time (h)↓	Params (M)↓	Model Memory (MB)↓	GPU Memory (MB)↓	Seen-Avg mAP ↑
LSTKC	2.5	28.65	109.55	10998	22.4
DKP	3.1	97.89	389.42	33518	22.9
DASK	9.5	29.75	113.94	11358	23.7
CORE [†] (Baseline)	3.6	57.30	219.10	20250	30.9
CKP	5.1	57.30	219.10	21402	42.2
CSDP	4.9	57.30	219.10	21402	46.9

and 1152MB of GPU memory (only a 5.7% increase). Moreover, our method introduces no additional parameters (Params) or model memory overhead compared to the CORE[†] baseline. Despite this, CSDP achieves significantly superior performance, with an improvement of **16.0%** (51.78% relative improvements) in Seen-Avg mAP under 30% random noise. These results demonstrate that our proposed methods achieve a more favorable trade-off between training overhead and performance.

Complexity analysis: Since only the backbone network is used during inference, the proposed method does not introduce any additional inference overhead. Regarding training, the additional computational cost primarily arises from the DBSCAN clustering algorithm. Specifically, given n samples and m identities, DBSCAN has a time complexity of $\mathcal{O}(n^2)$, while the remaining modules in our framework incur only linear complexity of $n + m$. As $n > m$ in practice, the overall time complexity is dominated by $\mathcal{O}(n^2)$.

In practice, however, DBSCAN is well-optimized in modern implementations, and its computational overhead is manageable. As a result, the overall training cost of our method is comparable to that of state-of-the-art approaches that introduce additional auxiliary network modules.

Ablations on Different Baselines: Our framework introduces a noisy data-aware correct knowledge accumulation paradigm that is compatible with various LReID approaches originally designed for clean data. In addition, it can be integrated with regularization-based methods such as CORE. To verify the effectiveness, as shown in Fig. 10, we adapt CKP and CSDP to multiple baselines, including LwF [69], LSTKC [11], LSTKC++ [38], DKP [12], DASK [37], the LNL method CORE,

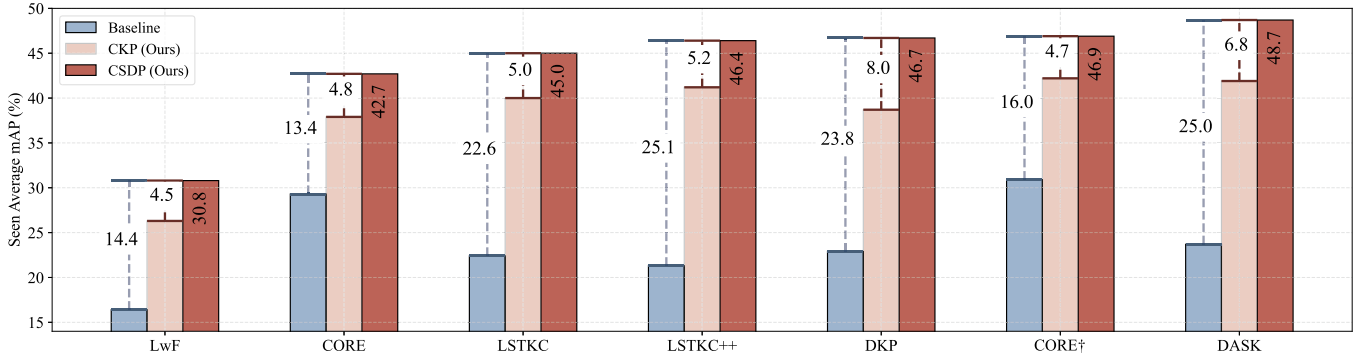


Fig. 10. Ablation on different baselines. The proposed CKP and CSDP can be readily integrated with existing LReID and LNL methods to significantly improve their Noisy LReID performance.

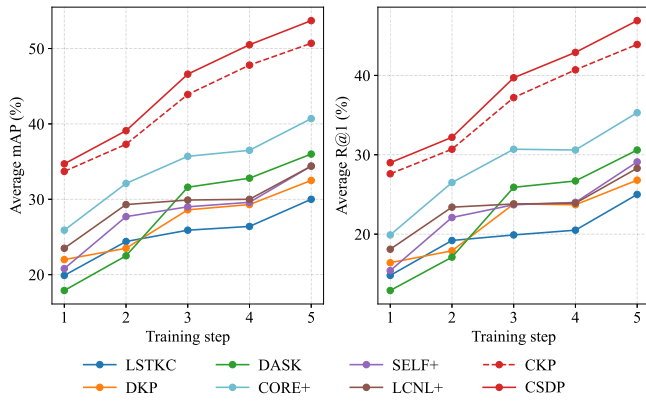


Fig. 11. Model performance tendency along the training steps.

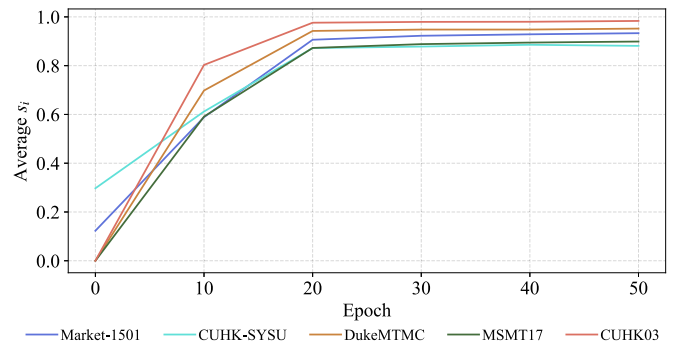


Fig. 13. Visualization of average s_i scores across all training samples during the training of each domain.

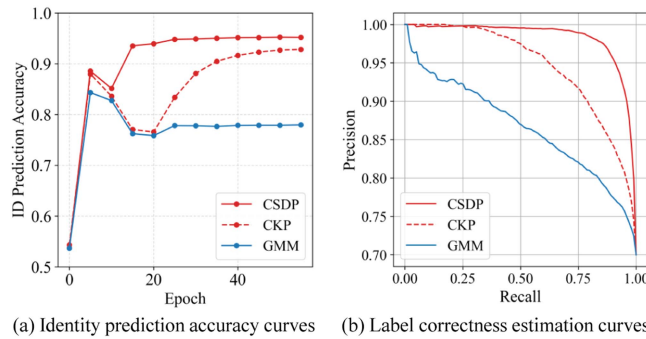


Fig. 12. Comparison of label purification capacity. (a) Identity prediction accuracy curves illustrate the accuracy of the predicted identity labels generated by our ILR module over training. (b) Label correctness estimation curves are the Precision–Recall curves of the estimated label confidence.

and the LNL+LReID method CORE[†] [14]. The results show that CKP and CSDP consistently outperform different baselines by large margins. Specifically, CSDP achieves **13.4%–25.1%** improvements across seven baselines and **4.5%–8.0%** gains over the conference version CKP. These improvements are attributed to the proposed local–global collaborative purification mechanism, which effectively alleviates the influence of erroneous labels. Furthermore, the self-paced learning and erroneous

knowledge filtering–driven noisy knowledge evolution scheme enhances learning stability, mitigates catastrophic remembering of noisy knowledge, and improves the robustness of the learned model.

Furthermore, we compare our approach with state-of-the-art replay-based continual learning methods and the results are reported in Table VI. The results show that CSDP consistently outperforms AER [86] and SPR [64], achieving improvements of at least **18.6%/23.7%** and **22.1%/21.1%** under Seen-Avg mAP/R@1 and Unseen-Avg mAP/R@1, respectively. These gains can be attributed to the fact that CSDP actively rectifies noisy labels, thereby improving the model’s ability to acquire reliable new knowledge.

We further evaluate a hybrid variant that integrates CSDP with AER [86]. Although CSDP+AER enhances anti-forgetting performance on the initial domain, its performance on subsequent domains degrades noticeably compared to CSDP alone. Overall, CSDP+AER yields **7.0%/8.8%** and **3.7%/3.1%** lower performance than CSDP under Seen-Avg mAP/R@1 and Unseen-Avg mAP/R@1. This degradation occurs because retaining historical exemplars during new-model training constrains the model’s adaptability to new data distributions, which in turn reduces the accuracy of data purification and limits performance in later domains. These results indicate that CSDP achieves a more favorable balance between knowledge acquisition and forgetting without relying on exemplar storage.

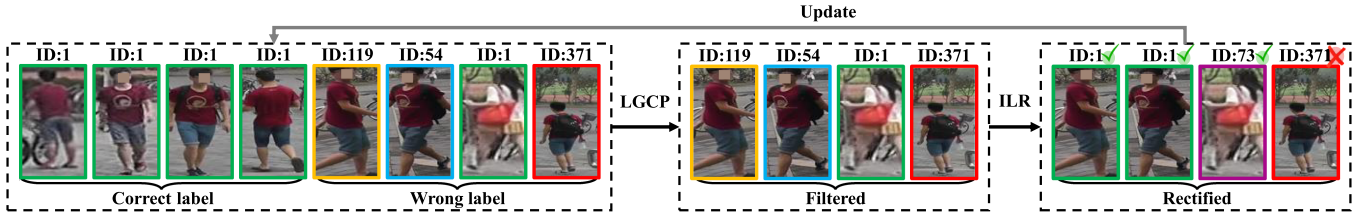


Fig. 14. Visualization of the erroneous label filtering and rectification. Each box color (green, orange, blue, red) indicates an identity label.

TABLE VI
ABLATION ON INCORPORATING OUR CSDP WITH REPLAY-BASED BASELINE

Method	Replay	Market-1501		CUHK-SYSU		DukeMTMC		MSMT17		CUHK03		Seen-Avg		Unseen-Avg	
		mAP	R@1	mAP	R@1	mAP	R@1	mAP	R@1	mAP	R@1	mAP	R@1	mAP	R@1
SPR [64]	✓	41.1	61.9	65.0	69.6	14.2	27.9	3.8	14.5	6.4	5.5	26.1	35.9	30.3	25.0
AER [86]	✓	46.5	69.2	68.5	71.6	16.1	27.4	3.3	9.6	7.2	6.3	28.3	36.8	31.6	25.8
CSDP	×	48.1	71.6	80.0	81.9	46.2	64.0	18.2	40.9	42.0	43.9	46.9	60.5	53.7	46.9
CSDP+AER [86]	✓	55.4	78.4	76.7	79.1	37.2	56.0	9.9	26.3	20.1	18.8	39.9	51.7	50.0	43.8

F. Visualization Results

Tendency of Generalizable Knowledge Accumulation: We visualize the average performance on unseen domains across lifelong training steps in Fig. 11. Compared with existing methods, both CKP and CSDP achieve higher performance starting from the first dataset and continue to demonstrate superior results as training progresses. This indicates that our method consistently consolidates correct knowledge while learning from noisy data. The improvement on the first dataset is mainly attributed to the robust learning capacity provided by the local relation-based purification mechanism. Subsequent increasing advantages stem from the erroneous knowledge filtering design, which effectively reduces the retention of misleading information. Moreover, CSDP consistently outperforms CKP across all domains, benefiting from the joint local-global knowledge utilization and the self-paced learning paradigm that enhances the model’s learning capacity.

Visualization of Label Purification Capacity: To assess the effectiveness of our data purification designs, we experimentally compare them against the Gaussian Mixture Model (GMM)-based strategy, which is widely adopted in noisy person re-identification works [3], [16], [17]. As shown in Fig. 12(a), all methods initially improve identity prediction accuracy in the early epochs. However, performance soon degrades due to overfitting to noisy data. Later, CKP and CSDP recover and achieve steady improvements, while the GMM-based method shows only marginal gains. Eventually, CKP and CSDP exceed GMM by more than 10% in accuracy, confirming the superiority of our local relation-based and local-global collaborative purification mechanisms.

In Fig. 12(b), we further present the Precision-Recall curves for label correctness estimation. The results show that CSDP achieves the best performance, followed by CKP, both significantly outperforming the GMM-based approach. The inferior performance of GMM arises from its reliance on distances to class centers, which is problematic in ReID scenarios where intra-identity variance is often large and inter-identity variance

small (see Fig. 2(c)). In contrast, CKP improves label correctness estimation by capturing local relations that better exploit inter-identity similarity. Furthermore, CSDP jointly leverages local and global knowledge, yielding superior label correctness estimation and thereby enhancing erroneous label filtering while retaining correct labels.

In Fig. 13, we additionally visualize the average s_i (3) over all samples within each domain across training epochs. The results reveal a **consistent upward trend** for our CSDP, indicating that the model progressively increases label confidence during training. This observation verifies that CSDP is able to continually purify noisy labels as learning proceeds.

Visualization of Erroneous Label Filtering and Rectification: As illustrated in Fig. 14, given a noisy dataset containing both correct and incorrect labels, the LGCP module leverages local and global cues to filter erroneous labels, while the correct labels are preserved for model training. Subsequently, the ILR module refines the remaining noisy labels by combining model predictions with label information. The rectified labels are then fed back to update the dataset, forming a loop that enables the model to gradually accumulate correct knowledge.

Although ILR can successfully rectify most incorrect labels, it may fail in extremely challenging cases, such as the example in the red box of Fig. 14, where the sample is assigned a wrong identity (371). This difficulty arises because the person appears in an abnormal position, making it hard for the model to extract discriminative features based on prior knowledge. Nevertheless, ILR remains valuable since it improves the utilization of wrongly labeled data, instead of discarding them outright.

Feature Distribution Visualization: To evaluate the training data utilization capacity, we visualize the feature distributions of different identities (represented by colors) in Fig. 15, comparing our method with existing approaches. Specifically, both LSTKC and CORE[†] exhibit significant inter-identity overlap due to their limited adaptivity to noisy labels. In contrast, CKP produces more cohesive clusters of clean and wrongly labeled samples, while CSDP achieves even clearer separation between identities.

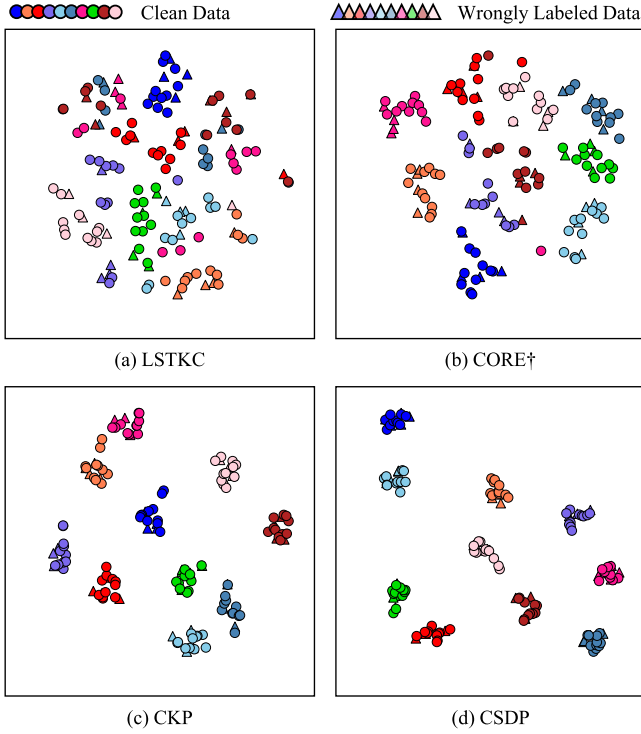


Fig. 15. Visualization of learned features in the training domain. Circle and triangle colors denote ground-truth identities, while triangle instances indicate samples assigned with erroneous labels during training. A more compact clustering within each color corresponds to higher-quality feature representations.

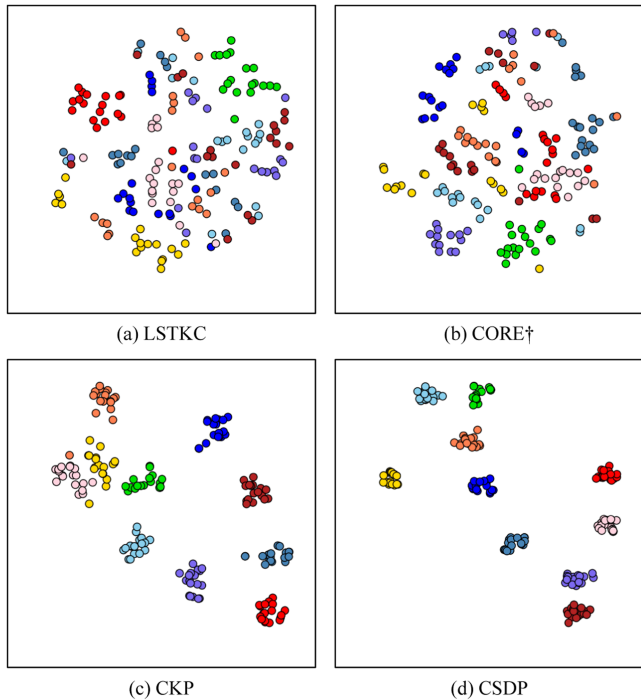


Fig. 16. Visualization of testing features. Each circle color denotes a ground-truth identity. More compact clustering within each color indicates higher-quality feature representations.

These results highlight the superiority of our method in mining correct inter-instance coherence under noisy label conditions.

We further visualize testing features in Fig. 16. The features extracted by CKP and CSDP form tighter and more distinct clusters than those produced by LSTKC and CORE[†]. This improvement stems from the proposed purification mechanisms and erroneous knowledge filtering, which enhance robustness against label noise. Moreover, the dual-knowledge utilization and self-paced learning paradigms further refine the feature space, yielding more accurate identity separation, as illustrated in Fig. 16(d).

VI. CONCLUSION

In this paper, we present a pioneering investigation into a practical and challenging task named Noisy Lifelong Person Re-Identification (Noisy LReID), which suffers not only from exacerbated catastrophic forgetting of correct knowledge but also from catastrophic remembering of erroneous knowledge. To facilitate future research in this area, we introduce a new benchmark, namely NLReID. Furthermore, we propose a novel and effective Continual Self-Paced Dual-Knowledge Purification (CSDP) framework, which collaboratively integrates a Local-Global Collaborative Purification mechanism, a Self-Paced Noisy Knowledge Evolution scheme, and an Iterative Label Rectification module. When used together, these components enable the purification of both new and historical knowledge. Extensive experimental results demonstrate that CSDP exhibits a superior capability in identifying and purifying noisy data compared with existing LReID, LNL, and their combined methods. While our study highlights that jointly leveraging local and global cues is an effective strategy for handling label noise in lifelong learning, such a process inevitably incurs additional computational and storage overhead. In the future, developing more efficient Noisy LReID frameworks remains a promising direction to further improve practical applicability.

ACKNOWLEDGMENT

The authors would like to acknowledge that this work was a research achievement of National Engineering Research Center of New Electronic Publishing Technologies.

REFERENCES

- [1] K. Xu, X. Zou, and J. Zhou, "LSTKC: Long short-term knowledge consolidation for lifelong person re-identification," in *Proc. AAAI Conf. Artif. Intell.*, 2024, vol. 38, no. 14, pp. 16202–16210.
- [2] Z. Cui, J. Zhou, and Y. Peng, "DMA: Dual modality-aware alignment for visible-infrared person re-identification," *IEEE Trans. Inf. Forensics Secur.*, vol. 19, pp. 2696–2708, 2024.
- [3] J. Shi et al., "Dual pseudo-labels interactive self-training for semi-supervised visible-infrared person re-identification," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2023, pp. 11218–11228.
- [4] V. D. Nguyen et al., "Tackling domain shifts in person re-identification: A survey and analysis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2024, pp. 4149–4159.
- [5] J. Shi et al., "Multi-memory matching for unsupervised visible-infrared person re-identification," in *Proc. Eur. Conf. Comput. Vis.*, 2024, pp. 456–474.
- [6] Y. Gong, Z. Zhong, Y. Qu, Z. Luo, R. Ji, and M. Jiang, "Cross-modality perturbation synergy attack for person re-identification," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2024, vol. 37, pp. 23352–23377.

- [7] Z. Cui, J. Zhou, Y. Peng, S. Zhang, and Y. Wang, "DCR-ReID: Deep component reconstruction for cloth-changing person re-identification," *IEEE Trans. Circuit Syst. Video Technol.*, vol. 33, no. 8, pp. 4415–4428, Aug. 2023.
- [8] G. Wu and S. Gong, "Generalising without forgetting for lifelong person re-identification," in *Proc. AAAI Conf. Artif. Intell.*, 2021, pp. 2889–2897.
- [9] N. Pu, W. Chen, Y. Liu, E. M. Bakker, and M. S. Lew, "Lifelong person re-identification via adaptive knowledge accumulation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 7897–7906.
- [10] K. Xu, F. Zhuo, J. Li, X. Zou, and J. Zhou, "Self-reinforcing prototype evolution with dual-knowledge cooperation for semi-supervised lifelong person re-identification," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2025, pp. 3564–3574.
- [11] Z. Sun and Y. Mu, "Patch-based knowledge distillation for lifelong person re-identification," in *Proc. ACM Int. Conf. Multimedia*, 2022, pp. 696–707.
- [12] K. Xu, X. Zou, Y. Peng, and J. Zhou, "Distribution-aware knowledge prototyping for non-exemplar lifelong person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2024, pp. 16604–16613.
- [13] Z. Cui, J. Zhou, and Y. Peng, "DKC: Differentiated knowledge consolidation for cloth-hybrid lifelong person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2025, pp. 3573–3582.
- [14] M. Ye, H. Li, B. Du, J. Shen, L. Shao, and S. C. Hoi, "Collaborative refining for person re-identification with label noise," *IEEE Trans. Image Process.*, vol. 31, pp. 379–391, 2022.
- [15] D. Fu et al., "Large-scale pre-training for person re-identification with noisy labels," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 2476–2486.
- [16] M. Yang, Z. Huang, P. Hu, T. Li, J. Lv, and X. Peng, "Learning with twin noisy labels for visible-infrared person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 14308–14317.
- [17] M. Yang, Z. Huang, and X. Peng, "Robust object re-identification with coupled noisy labels," *Int. J. Comput. Vis.*, vol. 132, pp. 2511–2529, 2024.
- [18] H. Zhao, T. Zhou, G. Long, J. Jiang, and C. Zhang, "Does continual learning equally forget all parameters?," in *Proc. Int. Conf. Mach. Learn.*, PMLR, 2023, pp. 42280–42303.
- [19] J. S. Smith, J. Tian, S. Halbe, Y.-C. Hsu, and Z. Kira, "A closer look at rehearsal-free continual learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 2409–2419.
- [20] Z. Cui, Y. Peng, X. Wang, M. Zhu, and J. Zhou, "Continual vision-language retrieval via dynamic knowledge rectification," in *Proc. AAAI Conf. Artif. Intell.*, 2024, vol. 38, no. 10, pp. 11704–11712.
- [21] W. Shi and M. Ye, "Prototype reminiscence and augmented asymmetric knowledge aggregation for non-exemplar class-incremental learning," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2023, pp. 1772–1781.
- [22] N.-R. Kim, J.-S. Lee, and J.-H. Lee, "Learning with structural labels for learning with noisy labels," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2024, pp. 27610–27620.
- [23] Z. Huang, J. Zhang, and H. Shan, "Twin contrastive learning with noisy labels," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 11661–11670.
- [24] W. Tang and T. M. Khoshgoftaar, "Noise identification with the k-means algorithm," in *Proc. IEEE Int. Conf. Tools Artif. Intell.*, 2004, pp. 373–378.
- [25] B. Nicholson, J. Zhang, V. S. Sheng, and Z. Wang, "Label noise correction methods," in *Proc. IEEE Int. Conf. Data Sci. Adv. Anal.*, 2015, pp. 1–9.
- [26] S. Li and X. Guo, "Distributed-clustering for data with heavy noise," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2018, vol. 31, pp. 7849–7857.
- [27] M. Ye, J. Shen, G. Lin, T. Xiang, L. Shao, and S. C. Hoi, "Deep learning for person re-identification: A survey and outlook," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 6, pp. 2872–2893, Jun. 2022.
- [28] S. He, H. Luo, P. Wang, F. Wang, H. Li, and W. Jiang, "TransReID: Transformer-based object re-identification," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2021, pp. 15013–15022.
- [29] Y. Fan, R. He, J. Liang, and B. Hu, "Self-paced learning: An implicit regularization perspective," in *Proc. AAAI Conf. Artif. Intell.*, 2017, vol. 31, no. 1, pp. 1877–1883.
- [30] K. Xu, H. Zhang, Y. Li, Y. Peng, and J. Zhou, "Mitigate catastrophic remembering via continual knowledge purification for noisy lifelong person re-identification," in *Proc. ACM Int. Conf. Multimedia*, 2024, pp. 5790–5799.
- [31] Z. Cui, J. Zhou, X. Wang, M. Zhu, and Y. Peng, "Learning continual compatible representation for re-indexing free lifelong person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2024, pp. 16614–16623.
- [32] J. Zhou, K. Xu, F. Zhuo, X. Zou, and Y. Peng, "Distribution-aware knowledge aligning and prototyping for non-exemplar lifelong person re-identification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 47, no. 12, pp. 10932–10948, Dec. 2025.
- [33] K. Xu, X. Zou, G. Hua, and J. Zhou, "Componential prompt-knowledge alignment for domain incremental learning," in *Proc. Int. Conf. Mach. Learn.*, 2025, pp. 70032–70046.
- [34] Z. Huang et al., "Lifelong unsupervised domain adaptive person re-identification with coordinated anti-forgetting and adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 14288–14297.
- [35] Z. Yang, D. Wu, W. Zhang, B. Li, and W. Wang, "Handling label uncertainty for camera incremental person re-identification," in *Proc. ACM Int. Conf. Multimedia*, 2023, pp. 6253–6263.
- [36] W. Ge et al., "Lifelong person re-identification by pseudo task knowledge preservation," in *Proc. AAAI Conf. Artif. Intell.*, 2022, pp. 688–696.
- [37] K. Xu, C. Jiang, P. Xiong, Y. Peng, and J. Zhou, "Dask: Distribution rehearsing via adaptive style kernel learning for exemplar-free lifelong person re-identification," in *Proc. AAAI Conf. Artif. Intell.*, 2025, vol. 39, no. 9, pp. 8915–8923.
- [38] K. Xu, Z. Liu, X. Zou, Y. Peng, and J. Zhou, "Long short-term knowledge decomposition and consolidation for lifelong person re-identification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 47, no. 9, pp. 7796–7811, Sep. 2025.
- [39] Q. Li, K. Xu, Y. Peng, and J. Zhou, "Exemplar-free lifelong person re-identification via prompt-guided adaptive knowledge consolidation," *Int. J. Comput. Vis.*, vol. 132, no. 11, pp. 4850–4865, 2024.
- [40] Y. Chen, M. Liu, X. Wang, F. Wang, A.-A. Liu, and Y. Wang, "Refining noisy labels with label reliability perception for person re-identification," *IEEE Trans. Multimedia*, vol. 25, pp. 9479–9490, 2023.
- [41] M. Liu, F. Wang, X. Wang, Y. Wang, and A. K. Roy-Chowdhury, "A two-stage noise-tolerant paradigm for label corrupted person re-identification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 7, pp. 4944–4956, Jul. 2024.
- [42] W. J. Yun, J. Park, and J. Kim, "Quantum multi-agent meta reinforcement learning," in *Proc. AAAI Conf. Artif. Intell.*, 2023, vol. 37, no. 9, pp. 11087–11095.
- [43] P. Kaushik, A. Kortylewski, A. Gain, and A. Yuille, "Understanding catastrophic forgetting and remembering in continual learning with optimal relevance mapping," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2021, pp. 1–22.
- [44] R. Zhao, W. Ouyang, and X. Wang, "Unsupervised salience learning for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2013, pp. 3586–3593.
- [45] C. Song, Y. Huang, W. Ouyang, and L. Wang, "Mask-guided contrastive attention model for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 1179–1188.
- [46] X. Teng, L. Lan, D. Chen, K. Xu, and N. Yin, "Relieving universal label noise for unsupervised visible-infrared person re-identification by inferring from neighbors," in *Proc. AAAI Conf. Artif. Intell.*, 2025, vol. 39, no. 7, pp. 7356–7364.
- [47] J. Zhang et al., "Badlabel: A robust perspective on evaluating and enhancing label-noise learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 6, pp. 4398–4409, Jun. 2024.
- [48] S. Ye, D. Chen, S. Han, and J. Liao, "Learning with noisy labels for robust point cloud segmentation," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2021, pp. 6443–6452.
- [49] S. Ye, D. Chen, S. Han, and J. Liao, "Robust point cloud segmentation with noisy annotations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 6, pp. 7696–7710, Jun. 2023.
- [50] D. Kim, K. Ryoo, H. Cho, and S. Kim, "SplitNet: Learnable clean-noisy label splitting for learning with noisy labels," *Int. J. Comput. Vis.*, vol. 133, no. 2, pp. 549–566, 2025.
- [51] M. Ye and P. C. Yuen, "PurifyNet: A robust person re-identification model with noisy labels," *IEEE Trans. Inf. Forensics Secur.*, vol. 15, pp. 2655–2666, 2020.
- [52] L. Yi et al., "When source-free domain adaptation meets learning with noisy labels," in *Proc. 11th Int. Conf. Learn. Representations*, 2023, pp. 1–33.
- [53] Y. Li, H. Han, S. Shan, and X. Chen, "Disc: Learning from noisy labels via dynamic instance-specific selection and correction," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 24070–24079.
- [54] J. Li, R. Socher, and S. C. Hoi, "DivideMix: Learning with noisy labels as semi-supervised learning," in *Proc. Int. Conf. Learn. Representations*, 2019, pp. 1–11.

- [55] D. Berthelot, N. Carlini, I. Goodfellow, N. Papernot, A. Oliver, and C. A. Raffel, "MixMatch: A holistic approach to semi-supervised learning," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2019, vol. 32, pp. 1–11.
- [56] Z. Zhu, Y. Song, and Y. Liu, "Clusterability as an alternative to anchor points when learning with noisy labels," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 12912–12923.
- [57] C. Xue, L. Yu, P. Chen, Q. Dou, and P.-A. Heng, "Robust medical image classification from noisy labeled data with global and local representation guided co-training," *IEEE Trans. Med. Imag.*, vol. 41, no. 6, pp. 1371–1382, Jun. 2022.
- [58] J. Shi et al., "Learning commonality, divergence and variety for unsupervised visible-infrared person re-identification," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2024, vol. 37, pp. 99715–99734.
- [59] X. Yin et al., "Robust pseudo-label learning with neighbor relation for unsupervised visible-infrared person re-identification," in *Proc. ACM Int. Conf. Multimedia*, 2024, pp. 2242–2251.
- [60] S. Li, X. Xia, J. Deng, S. Ge, and T. Liu, "Transferring annotator-and instance-dependent transition matrix for learning from crowds," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 11, pp. 7377–7391, Nov. 2024.
- [61] Y. Liu, H. Cheng, and K. Zhang, "Identifiability of label noise transition matrix," in *Proc. Int. Conf. Mach. Learn.*, 2023, pp. 21475–21496.
- [62] D. Cheng et al., "Class-dependent label-noise learning with cycle-consistency regularization," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2022, vol. 35, pp. 11104–11116.
- [63] R. Das and S. Sanghavi, "Understanding self-distillation in the presence of label noise," in *Proc. Int. Conf. Mach. Learn.*, 2023, pp. 7102–7140.
- [64] C. D. Kim, J. Jeong, S. Moon, and G. Kim, "Continual learning on noisy data streams via self-purified replay," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2021, pp. 537–547.
- [65] N. Karim, U. Khalid, A. Esmaili, and N. Rahnavard, "Cnll: A semi-supervised approach for continual noisy label learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 3878–3888.
- [66] G. Guo, Z. Wei, and J. Cheng, "Enhancing continual noisy label learning with uncertainty-based sample selection and feature enhancement," in *Proc. Chin. Conf. Pattern Recognit. Comput. Vis.*, 2023, pp. 498–510.
- [67] G. Li, P. Wang, Q. Luo, Y. Liu, and W. Ke, "Online noisy continual relation learning," in *Proc. AAAI Conf. Artif. Intell.*, 2023, vol. 37, no. 11, pp. 13059–13066.
- [68] E. Schubert, J. Sander, M. Ester, H. P. Kriegel, and X. Xu, "DbSCAN revisited, revisited: Why and how you should (STILL) use DBSCAN," *ACM Trans. Database Syst.*, vol. 42, no. 3, pp. 1–21, 2017.
- [69] Z. Li and D. Hoiem, "Learning without forgetting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 12, pp. 2935–2947, Dec. 2018.
- [70] Z. Gao et al., "Maintaining fairness in logit-based knowledge distillation for class-incremental learning," in *Proc. AAAI Conf. Artif. Intell.*, vol. 39, no. 16, 2025, pp. 16763–16771.
- [71] J. Dong et al., "LLM-assisted entropy-based adaptive distillation for unsupervised fine-grained visual representation learning," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2025, pp. 383–392.
- [72] F. Zhu, X.-Y. Zhang, Z. Cheng, and C.-L. Liu, "PASS++: A dual bias reduction framework for non-exemplar class-incremental learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 47, no. 8, pp. 7123–7139, Aug. 2025.
- [73] C. Yu, Y. Shi, Z. Liu, S. Gao, and J. Wang, "Lifelong person re-identification via knowledge refreshing and consolidation," in *Proc. AAAI Conf. Artif. Intell.*, 2023, pp. 3295–3303.
- [74] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable person re-identification: A benchmark," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1116–1124.
- [75] E. Ristani, F. Solera, R. Zou, R. Cucchiara, and C. Tomasi, "Performance measures and a data set for multi-target, multi-camera tracking," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 17–35.
- [76] T. Xiao, S. Li, B. Wang, L. Lin, and X. Wang, "End-to-end deep learning for person search," 2016, *arXiv:1604.01850*.
- [77] L. Wei, S. Zhang, W. Gao, and Q. Tian, "Person transfer GAN to bridge domain gap for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 79–88.
- [78] W. Li, R. Zhao, T. Xiao, and X. Wang, "DeepReID: Deep filter pairing neural network for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 152–159.
- [79] W. Li, R. Zhao, and X. Wang, "Human reidentification with transferred metric learning," in *Proc. Asian. Conf. Comput. Vis.*, 2012, pp. 31–44.
- [80] W. Li and X. Wang, "Locally aligned feature transforms across views," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2013, pp. 3594–3601.
- [81] D. Gray and H. Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features," in *Proc. Eur. Conf. Comput. Vis.*, 2008, pp. 262–275.
- [82] M. Hirzer, C. Belezai, P. M. Roth, and H. Bischof, "Person re-identification by descriptive and discriminative classification," in *Image Analysis*. Berlin, Germany: , 2011, pp. 91–102.
- [83] Home Office Scientific Development Branch, "Imagery library for intelligent detection systems (i-LIDS)," in *Proc. IET Conf. Crime Secur.*, 2006, pp. 445–448.
- [84] C. C. Loy, T. Xiang, and S. Gong, "Time-delayed correlation analysis for multi-camera activity understanding," *Int. J. Comput. Vis.*, vol. 90, no. 1, pp. 106–129, 2010.
- [85] H. Zhao et al., "Spindle Net: Person re-identification with human body region guided feature decomposition and fusion," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 907–915.
- [86] M. Millunzi et al., "May the forgetting be with you: Alternate replay for learning with noisy labels," in *Proc. Brit. Mach. Vis. Conf.*, 2024, pp. 1–16.
- [87] R. J. Campello, D. Moulavi, and J. Sander, "Density-based clustering based on hierarchical density estimates," in *Proc. Pacific-Asia Conf. Knowl. Discov. Data Mining*, 2013, pp. 160–172.
- [88] B. Nadler and M. Galun, "Fundamental limitations of spectral clustering," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, vol. 19, 2006, pp. 1017–1024.



Kunlun Xu received the BE and M.S. degrees from the Huazhong University of Science and Technology, in 2020 and 2023, respectively. He is currently working toward the PhD degree with the Wangxuan Institute of Computer Technology, Peking University. He has authored 15 papers in international journals and conferences, including *IEEE Transactions on Pattern Analysis and Machine Intelligence* (IEEE TPAMI), *IJCV*, *NeurIPS*, *ICML*, *CVPR*, *ICCV*, *AAAI*, and *ACM MM*. His research interests include computer vision, lifelong learning, and person re-identification.

He is also a regular reviewer member for a number of journals and conferences, e.g., *IEEE TPAMI*, *IEEE Transactions on Image Processing*, *IEEE Transactions on Multimedia*, *ICML*, *NeurIPS*, *ICLR*, *CVPR*, *ICCV*, *AAAI*, and *ACM MM*.



Jiangmeng Li received the MS degree from New York University, New York, USA, in 2018, and the PhD degree from the University of Chinese Academy of Sciences, Beijing, China, in 2025. He is currently an associate professor (tenure-track) with the Institute of Software, Chinese Academy of Sciences. He has authored or coauthored more than 50 papers in journals and conferences, such as *IEEE Transactions on Knowledge and Data Engineering* (TKDE), *International Journal of Computer Vision* (IJCV), *International Conference on Machine Learning* (ICML),

and Conference and Workshop on Neural Information Processing Systems (NeurIPS). His research interests include multi-modal learning, self-supervised learning, and graph learning.



Yufei Guo received the BS and PhD degrees from the Beijing Institute of Technology and Peking University, in 2015 and 2020, respectively. He is currently a senior engineer with Intelligent Science & Technology Academy, CASIC, Beijing, China. His interests include neuromorphic computing, model compression, data generation, and LLM (large language model).



Yuxin Peng (Fellow, IEEE) received the PhD degree in computer application technology from Peking University, Beijing, China, in 2003. He was the recipient of the National Science Fund for Distinguished Young Scholars of China. He is currently the boya distinguished professor with the Wangxuan Institute of Computer Technology, Peking University. He has authored over 200 papers, including more than 100 papers in top-tier journals and conference proceedings. He has submitted 51 patent applications and has been granted 39 of them. His research interests mainly

include cross-media analysis and reasoning, image and video recognition and understanding, and computer vision. He was the recipient of the Best Paper Award at MMM 2019 and NCIG 2018, the First Prize of the Beijing Science and Technology Award in 2016 (ranking first), and First Prize of the Scientific and Technological Progress Award of the Chinese Institute of Electronics in 2020 (ranking first). He also led his team to win the First Place in the video semantic search evaluation of TRECVID ten times in recent years. He is also the associate editor for *IEEE Transactions on Multimedia* and *IEEE Transactions on Circuits and Systems for Video Technology*.



Jiahuan Zhou received the BE degree from Tsinghua University, in 2023, and the PhD degree from the Department of Electrical Engineering & Computer Science, Northwestern University, in 2018. He is currently a tenure-track assistant professor with the Wangxuan Institute of Computer Technology, Peking University. He has authored more than 60 papers in international journals and conferences including *Nature Communications*, *Nature Synthesis*, *IEEE Transactions on Pattern Analysis and Machine Intelligence* (IEEE T-PAMI), *IJCV*, *IEEE Transactions on Image*

Processing (IEEE TIP), *IEEE Transactions on Information Forensics and Security*, *CVPR*, *ICCV*, *ECCV*, *ICML*, *NeurIPS*, *ICLR*, *AAAI*, *IJCAI*, and *ACM MM*. His research interests include computer vision, deep learning, and machine learning. He is also an area chair for *CVPR*, *ICML*, *NeurIPS*, *ICLR*, *ICME*, and *ICPR*, an associate editor for the *Springer Journal of Machine Vision and Applications* (MVA), a regular reviewer member for several journals and conferences, e.g., T-PAMI, IJCV, TIP, CVPR, ICCV, ECCV, NeurIPS, ICML, ICLR, and AAAI.