Active Learning for Animal Re-Identification with Ambiguity-Aware Sampling

Depanshu Sani, Mehar Khurana, Saket Anand

Indraprastha Institute of Information Technology, Delhi, India
{depanshus, mehar21541, anands}@iiitd.ac.in
https://sites.google.com/iiitd.ac.in/aas/

Abstract

Animal re-identification (Re-ID) has recently gained substantial attention in the AI research community due to its high impact on biodiversity monitoring and unique research challenges arising from environmental factors. The subtle distinguishing patterns like stripes or spots, handling new species and the inherent open-set nature make the problem even harder. To address these complexities, foundation models trained on labeled, large-scale and multi-species animal Re-ID datasets have recently been introduced to enable zero-shot Re-ID. However, our benchmarking reveals significant gaps in their zero-shot Re-ID performance for both known and unknown species. While this highlights the need for collecting labeled data in new domains, exhaustive annotation for Re-ID is laborious and requires domain expertise. Our analyses also show that existing unsupervised (USL) and active learning (AL) Re-ID methods underperform for animal Re-ID. To address these limitations, we introduce a novel AL Re-ID framework that leverages complementary clustering methods to uncover and target structurally ambiguous regions in the embedding space for mining pairs of samples that are both informative and broadly representative. Oracle feedback on these pairs, in the form of must-link and cannot-link constraints, facilitates a simple annotation interface, which naturally integrates with existing USL methods through our proposed constrained clustering refinement algorithm. Through extensive experiments, we demonstrate that, by utilizing only 0.033% of all possible annotations, our approach consistently outperforms existing foundational, USL and AL baselines. Specifically, we report an average improvement of 10.49%, 11.19% and 3.99% (mAP) on 13 wildlife datasets over foundational, USL and AL methods, respectively, while attaining state-of-the-art performance on each dataset. Furthermore, we also show an improvement of 11.09%, 8.2% and 2.06% (AUC ROC) for unknown individuals in an open-world setting. We also present results on 2 publicly available person Re-ID datasets, showing average gains of 7.96% and 2.86% (mAP) over existing USL and AL Re-ID methods.

1 Introduction

Re-ID is a critical task that aims to match instances of the same individual across different images. While traditionally explored in the context of human surveillance, Re-ID has found growing relevance in wildlife monitoring, where it is used to track individual animals over time. The ability to Re-ID individuals plays a pivotal role in ecological studies, population estimation, behavioral tracking and conservation planning. Non-invasive approaches, like the use of camera trap images, often require manual processing to ensure reliable re-identification.

Animal Re-ID has recently gained substantial attention in the AI research community, not only due to its high impact on biodiversity monitoring, but also because of its unique and significantly more complex challenges compared to other Re-ID tasks. For instance, unlike human Re-ID where the key distinguishing attributes are same for all, animal Re-ID depends on visual patterns that are *species-dependent*, such as stripes for tigers, spots for cheetahs and rosettes for leopards. To address this, several AI-enabled systems have been developed for species-agnostic animal Re-ID (Jiao et al. 2023; Čermák et al. 2024; Otarashvili et al. 2024). At their core, these systems leverage foundation models that are trained on large-scale multi-species wildlife datasets that are manually annotated to find instances of the same individual across diverse species and geographies.

Despite recent advances, existing foundation models for animal Re-ID (Čermák et al. 2024; Otarashvili et al. 2024) struggle to perform adequately for new environments. Moreover, adapting Re-ID systems to new environments remains challenging due to the high cost of manual annotation. Manually identifying individuals in a large image set is tedious and error-prone as it needs discerning fine-grained patterns in images. A natural way to approach this is to frame supervision in terms of pairwise annotations, where annotators simply indicate whether two images are of the same individual. Unfortunately, this labeling process is expensive and time-consuming as it requires identifying matching pairs within a large dataset, where the number of comparisons grows quickly with the dataset size.

To mitigate the cost of exhaustive annotation, many recent unsupervised Re-ID approaches (Ge et al. 2020; Zhang et al. 2021; Dai et al. 2021; Wang et al. 2021; Wu et al. 2021; Zou et al. 2023; Yin et al. 2023) rely on pseudo-labeling via clustering, where the model is iteratively refined using automatically generated labels from the structure of the feature space. However, the effectiveness of pseudo-labeling hinges critically on cluster purity. Errors in clustering can introduce harmful supervision. Broadly, clustering errors can be cate-

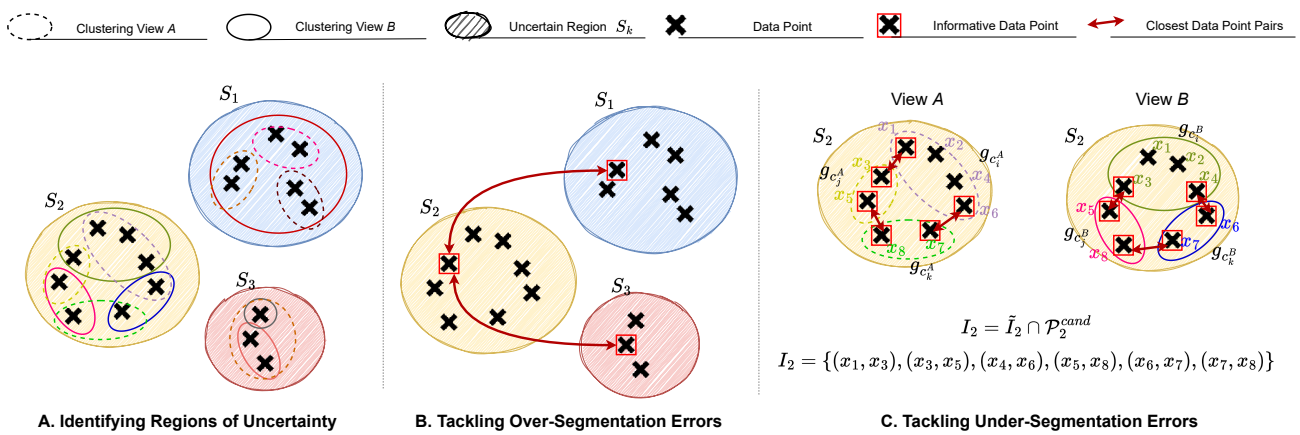


Figure 1: An illustration of Ambiguity-Aware Sampling (AAS). **(A.) Identifying Regions of Uncertainty.** The disagreements between the clustering views A and B are used to identify regions of uncertainty, $S = \{S_1, S_2, S_3\}$, based on the transitive closure of *partially overlapping* clusters from A and B. **(B.) Resolving Over-Segmentation Errors.** The medoid $\mathbf{m}_k$ is chosen to be the representative of an uncertain region S_k. For each medoid $\mathbf{m}_k$, we obtain its nearest k_{max} neighboring medoids, $\mathbf{m}_{k'}$ $\forall k' \neq k$, having similarity $> s_{min}$ to form most informative image pairs. **(C.) Resolving Under-Segmentation Errors.** $\tilde{I}_k$ is the set of all pairs that are in disagreement within S_k. Note that $\tilde{I}_k$ might contain redundant pairs that do not provide additional value if annotated. For instance, pairs (x_1, x_6), (x_2, x_6) and (x_4, x_6) in the sub-figure represent the same ambiguity, i.e. whether the group of consistently clustered images $\{x_1, x_2, x_4\}$ should be clustered with $\{x_6\}$. Hence, to filter out redundant pairs, we define $\mathcal{P}_k^{cand}$ to be a set of candidate image pairs that contains the nearest inter-cluster neighbors for both the methods, which may or may not be inconsistent. Therefore, for each uncertain region S_k, we define the set of most informative and non-redundant *inconsistent* image pairs as $I_k = \tilde{I}_k \cap \mathcal{P}_k^{cand}$. Finally, the set of informative medoid pairs ($\mathcal{U}_{os}$) and non-redundant inconsistent pairs ($\mathcal{U}_{us}$) defines our sampling pool $\mathcal{U} = \{\mathcal{U}_{os} \cup \mathcal{U}_{us}\}$ from which the image pairs are sampled for annotation.

gorized into two types: *under-segmentation*, where samples from different identities are erroneously grouped into the same cluster, and *over-segmentation*, where samples from the same identity are incorrectly split into multiple clusters. Both scenarios are detrimental—under-segmentation leads to identity confusion and suppresses inter-class separability, while over-segmentation fragments identity representations and weakens intra-class consistency. These challenges highlight the fragility of unsupervised pseudo-labeling and underscore the importance of guiding the learning process using reliable supervisory signals.

This motivates the use of active learning (AL), which seeks to identify and label the most informative examples under a constrained annotation budget. In the Re-ID setting, where supervision is naturally framed through pairwise comparisons, not all image pairs are equally valuable. Most pairs offer little learning signal—being either obviously matched or mismatched. Instead, the greatest benefit lies in annotating ambiguous pairs, especially those near decision boundaries. Moreover, to ensure robust generalization and avoid overfitting, it is equally important that selected pairs are diverse, covering a wide range of identities and appearance variations. Traditional works explored AL in the context of Re-ID using an entropy-based (Das, Panda, and Roy-Chowdhury 2015), rank-based (Wang et al. 2018) and reinforcement learning-based (Liu et al. 2019; Joshi and Subramanyam 2022) selection criterion. The primary focus of many existing solutions has been on person or vehicle Re-ID, often resulting in works that also target modeling the multi-camera setup (Das, Chakraborty, and Roy-Chowdhury 2014; Lin et al. 2017; Roy et al. 2018). With advances in unsupervised deep clustering methods, recent methods exploit cluster statistics to identify the most informative pairs (Hu et al. 2021) or triplets (Han et al. 2021; He et al. 2024). While NIS (Perez et al. 2024) designs an AL strategy to estimate the population size for wildlife datasets using nested information sampling, to the best of our knowledge, there exists no prior work that leverages AL for animal Re-ID.

In this work, we argue that the pseudo-labels obtained via different clustering algorithms are often structurally distinct, reflecting their differing inductive biases. Therefore, the variations in these pseudo labels can reveal ambiguities within the feature space. By identifying the disagreements between these complementary clustering views, we aim to uncover the most uncertain image pairs that can help reduce *under-segmentation* errors. Conversely, agreement between methods does not guarantee correctness, as it can hide *over-segmentation* errors. Therefore, we propose a novel active sampling strategy which can be used to sample pairs for addressing both under- and over-segmentation (Fig 1). Moreover, our AL Re-ID framework is designed in a way that can be seamlessly integrated with existing unsupervised or semi-supervised deep clustering methods for efficient training. To enable this, we introduce a *post-hoc* constrained clustering algorithm that can be used to incorporate the human feedback into the pseudo labels generated via the base clustering method. Our contributions are summarized below:

1. We propose Ambiguity-Aware Sampling (AAS) strategy that leverages disagreements between two complementary clustering algorithms—DBSCAN (density-based) and FINCH (nearest-neighbor-based)—to identify and sample most informative, uncertain and diverse pairs of images for annotation.
2. We introduce a Non-Parametric, Plug-and-Play (NP3) algorithm, that is a *post-hoc* constrained clustering method and is agnostic to the underlying clustering technique used to obtain the labels.

By integrating these strategies into the AL Re-ID pipeline, we demonstrate the following:

1. Existing foundation models, unsupervised training strategies and AL Re-ID methods show limited gains over a pre-trained ResNet-50 baseline on animal Re-ID benchmarks.
2. NP3 algorithm enables seamless integration of AL Re-ID techniques with existing unsupervised training pipelines.
3. State-of-the-art performance on 13 wildlife and 2 person Re-ID benchmark datasets.

2 Related Work

Animal ReID: Ecologists have employed various Re-ID methods for decades (Schneider et al. 2018; Tuia et al. 2022), including tagging (Wolcott 1995; Block et al. 2001; Cooke et al. 2004), scarring (Hammond, Mizroch, and Donovan 1990; Bertulli, Rasmussen, and Rosso 2016), and banding (Carroll et al. 2017). However, these methods are labor-intensive, expensive (Schneider et al. 2018), and can cause pain or impact animal behavior (Walker et al. 2012).

Due to the active shift towards non-invasive monitoring, camera trap images powered by computer vision techniques have emerged as an efficient and scalable alternative. Early methods relied on classical techniques like SIFT (Ravela et al. 2014; Bolger et al. 2012; Crall et al. 2013), but the field has advanced significantly with deep learning (Shukla et al. 2019a,b). A substantial body of research is focused on curating large-scale benchmark datasets that facilitate the design of modern algorithms (Guo et al. 2020; Lu et al. 2023; Čermák et al. 2024; Adam et al. 2025). At the same time, researchers have also focused on developing deep learning models for multi-species animal Re-ID. UniReID (Jiao et al. 2023) uses pre-trained CLIP as the backbone model and learns the visual and textual guidance generators to obtain semantic knowledge, which is then used to guide their UniReID focus on salient visual content via the text-guided attentive module. MegaDescriptor (Čermák et al. 2024) leverages the SWIN transformer architecture (Liu et al. 2021) and is trained on 2.8M images spanning 30K identities. MiewID (Otarashvili et al. 2024) is another foundation model that utilizes a simpler convolutional backbone and is trained on 2.3M images of 37K individuals. Despite the availability of such foundation models, we observe that they often underperform on unseen data, highlighting a crucial need for fine-tuning. While WildFusion (Cermak et al. 2025) recently proposed fusing and calibrating global (e.g. MegaDescriptor) and local (e.g. SIFT) feature matching scores, we believe fine-tuning Re-ID models on unseen data is still crucial.

Unsupervised ReID: These approaches are broadly categorized as either unsupervised domain adaptive (UDA) or fully unsupervised learning (USL). A key difference is that USL methods operate on unlabeled data without relying on a source-domain dataset. SpCL (Ge et al. 2020) is a pioneering example for both UDA and USL that uses a self-paced contrastive learning framework with a hybrid memory. This hybrid memory is crucial for generating the necessary signals to learn effective features by jointly distinguishing between source classes, target clusters and unclustered instances. Subsequent research has built on this foundation by focusing on improving cluster proxies and memory mechanisms (Dai et al. 2021; Wang et al. 2021; Wu et al. 2021; Zou et al. 2023; Yin et al. 2023), dynamically refining pseudo-labels (Zhang et al. 2021; Zheng et al. 2021) and leveraging multi-camera information to boost performance (Das, Chakraborty, and Roy-Chowdhury 2014; Lin et al. 2017; Wang et al. 2021; Li et al. 2023). Given its inherent generalizability, we use the SpCL framework as our base training method, acknowledging that more recent USL methods could similarly be integrated.

AL for ReID: Several recent efforts have explored AL strategies, tailored to the Re-ID setting, each introducing specific heuristics to select diverse informative image pairs (Das, Panda, and Roy-Chowdhury 2015; Wang et al. 2018; Liu et al. 2019; Joshi and Subramanyam 2022; Das, Chakraborty, and Roy-Chowdhury 2014; Lin et al. 2017; Roy et al. 2018; Han et al. 2021). For instance, the MASS framework (Hu et al. 2021) focuses on identifying and purifying impure clusters by forming coarse clusters using FINCH (Sarfraz, Sharma, and Stiefelhagen 2019) and then querying annotations for a central pivot of the coarse cluster and its neighbors. While effective in moderately clean settings, such strategies can waste annotation budget on highly under-segmented clusters containing multiple identities, where refinement is unlikely to succeed. The state-of-the-art in AL for person Re-ID, LBAS (He et al. 2024), aims to balance positive and negative supervision within the annotation budget. They first generate clusters using DBSCAN (Ester et al. 1996), then mine informative triplets within clusters by selecting anchor-positive-negative sets using farthest point sampling. Although this approach promotes diversity in the labeled pairs, it is highly sensitive to DBSCAN's parameters and tends to discard the detected outliers, which often represent the most ambiguous and thus most informative samples. We note that none of the recent AL Re-ID methods have open-sourced their implementation. Moreover, to the best of our knowledge, there exists no AL Re-ID method that is tested on wildlife datasets. The closest and the most relevant work to ours is NIS (Perez et al. 2024), which designs an AL technique based on nested importance sampling to estimate the size of the animal population in the wildlife datasets. Specifically, it constructs a fully connected similarity graph and samples vertices with estimated low degrees, followed by sampling their most similar neighbors. However, it treats all sampled vertices as equally

uncertain and uniformly allocates a fixed number of annotations per sampled vertex, leading to inefficient labeling when the ambiguity is highly variable.

3 Methodology

Let $\mathcal{D} = \{x_1, \ldots, x_n\}$ be the dataset comprising of cropped images of individuals. Assume $f : \mathcal{D} \to \mathbb{R}^d$ be a feature extractor that maps an image x_i to its vector representation $\mathbf{x}_i$. We employ two clustering methods, A and B, on all the feature vectors $\mathcal{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ to get clusters $\mathcal{C}^A = \{c_1^A, \ldots, c_{|\mathcal{C}^A|}^A\}$ and $\mathcal{C}^B = \{c_1^B, \ldots, c_{|\mathcal{C}^B|}^B\}$, respectively. Each image $x_k \in \mathcal{D}$ is assigned a pseudo-identity $\hat{y}_k \in \mathcal{C}^A$ by method A and $\tilde{y}_k \in \mathcal{C}^B$ by method B. We denote the set of images belonging to a specific cluster c_i^A as $\mathcal{Y}_{c_i^A}$, and similarly for c_j^B as $\mathcal{Y}_{c_j^B}$.

3.1 Framework for Ambiguity-Aware Sampling

Our goal is to create a pool of unlabeled image pairs ($\mathcal{U}$), which is a subset of all image pairs ($\mathcal{D} \times \mathcal{D}$). This pool $\mathcal{U}$ is designed to contain only the most informative pairs, identified through an analysis of agreements and disagreements between the complementary clustering methods A and B. Due to page length restrictions, we provide a brief overview of the process for constructing $\mathcal{U}$ and defining a marginal distribution over its image pairs, which subsequently guides the sampling of pairs for human annotation. Our AL Re-ID framework is summarized in Algorithm 1. A detailed explanation of each step is provided in the supplementary[1].

A. Identifying Regions of Uncertainty via Disagreement: We focus on pairs of clusters with a partial overlap ($0 < IoU < 1$), which indicate disagreements between methods A and B. We define a set of connected components, $S = \{S_1, \ldots, S_M\}$, where each S_k is a distinct "region of uncertainty". These regions are the transitive closure of partially overlapping clusters, and their images are "tangled" as their assignments are inconsistent across the two methods (See Fig. 1A). This uncertainty primarily signals potential under-segmentation errors, where one method's cluster should be further divided based on the other's assignments. Conversely, these errors can also be viewed as over-segmentation errors for the other clustering method.

To address both under- and over-segmentation issues, our AL approach builds a sampling pool $\mathcal{U}$ containing diverse pairs both from *within* (to resolve under-segmentation) and *across* (to resolve over-segmentation) the identified regions of uncertainty. A key benefit of this strategy is that if the size of $\mathcal{U}$ is smaller than the annotation budget, our method can reduce labeling effort by only querying the most informative pairs. In cases where there are no disagreements, we do not actively sample and instead rely on the base USL training.

B. Tackling Over-Segmentation Errors: To address over-segmentation errors, where a single identity is split across multiple uncertain regions, we construct a set of informative pairs, $\mathcal{U}_{os}$. Our goal is to find potential must-links between images in different regions that may belong to the same individual. Our strategy samples representative pairs based on the medoids of these regions of uncertainty. For each region S_k, its medoid is the image whose feature vector is closest to all other feature vectors within that region. From the collection of all medoids $\{\mathbf{m}_1, \ldots, \mathbf{m}_M\}$, we generate candidate pairs for $\mathcal{U}_{os}$ by identifying the $k_{\max}$ nearest neighbors for each medoid in the feature space, as shown in Fig. 1(B). A pair of images is included in $\mathcal{U}_{os}$ only if their similarity is greater than or equal to a predefined threshold, $s_{\min}$. This threshold is crucial for filtering out obvious mismatches and focusing on pairs that are potentially from the same identity.

C. Tackling Under-Segmentation Errors: To identify informative pairs for resolving under-segmentation errors, we focus on fine-grained disagreements between clustering methods A and B. Our goal is to find image pairs that can resolve whether they should be assigned to different clusters.

First, we define $\tilde{I}_k$ as the set of all inconsistent pairs in a region S_k, using the symmetric difference ($\triangle$) between the pairs grouped by method A (A_k^+) and method B (B_k^+):

$$\tilde{I}_k = A_k^+ \triangle B_k^+ \tag{1}$$

This set, however, can contain redundant pairs that resolve the same ambiguity. To create a more efficient set of queries, we propose resolving each ambiguity by querying only the single, most informative pair. The most ambiguous boundaries are represented by the "closest pairs" between distinct clusters. We define $\mathcal{P}_k^{\text{cand}}$ as the set of these closest pairs for all distinct cluster pairs within S_k (for both A and B):

$$\mathcal{P}_k^{\text{cand}} = \left\{ \underset{(x_u, x_v)}{\arg\min} \left(\text{dist}(x_u, x_v)\right) \mid (x_u, x_v) \in \mathcal{Y}_{c_i^\zeta} \times \mathcal{Y}_{c_j^\zeta} \right. \\ \left. \forall c_i^\zeta, c_j^\zeta \in S_k, i \neq j \text{ and } \zeta \in \{A, B\} \right\} \tag{2}$$

We then filter the set of initial inconsistent pairs, $\tilde{I}_k$, by intersecting it with our set of closest candidate pairs (Fig. 1(C)):

$$I_k = \tilde{I}_k \cap \mathcal{P}_k^{\text{cand}} \quad \forall\, k \in \{1, \ldots, M\} \tag{3}$$

Finally, the sampling pool for under-segmentation, $\mathcal{U}_{us}$, is the union of all I_k across all uncertainty regions, comprising the most informative and non-redundant pairs.

D. Defining Marginal Probability Distribution over $\mathcal{U}$: Our objective is to define the marginal probability distribution $P(Y)$ over any pair of images $Y = (x_u, x_v)$ in the total sampling pool $\mathcal{U} = \mathcal{U}_{os} \cup \mathcal{U}_{us}$. This distribution is a weighted combination of our strategies for resolving over- and under-segmentation errors, with a prior probability $\epsilon \in [0, 1]$ controlling their relative importance.

For pairs from the over-segmentation pool $\mathcal{U}_{os}$, the conditional probability is their feature similarity:

$$P(Y = (x_u, x_v) \mid Y \in \mathcal{U}_{os}) = \frac{\text{sim}(x_u, x_v)}{\sum \text{sim}(x'_u, x'_v)} \tag{4}$$

where the sum in the denominator runs $\forall (x'_u, x'_v) \in \mathcal{U}_{os}$. For pairs from the under-segmentation pool $\mathcal{U}_{us}$, we recognize that the relevance of an inconsistent pair depends on its relationship—whether it's an *inlier–inlier* (β_1), *inlier–outlier* (β_2) or *outlier–outlier* (β_3) pair, as identified by DBSCAN. We model the conditional probability for these pairs as a

[1] **Extended Version:** https://arxiv.org/abs/2511.06658

product of three factors: a prior probability (π_i) for the relationship type, a likelihood ($\rho_{k|i}$) of sampling from the uncertain region and a conditional probability ($\omega_{y|ik}$) of sampling the specific pair itself. This approach prioritizes regions with more ambiguous underlying structures. These probabilistic components are formally defined in the supplementary. The conditional probability of sampling a pair from $\mathcal{U}_{us}$ is:

$$P(Y = (x_u, x_v) \mid Y \in \mathcal{U}_{us}) = \pi_i \cdot \rho_{k|i} \cdot \omega_{y|ik} \quad (5)$$

The final marginal probability distribution for $Y \in \mathcal{U}$ is:

$$P(Y) = \epsilon \cdot P(Y \mid Y \in \mathcal{U}_{os}) + (1-\epsilon) \cdot P(Y \mid Y \in \mathcal{U}_{us}) \quad (6)$$

E. Refining Pseudo Labels with Pairwise Constraints: To address the limitations of existing constrained clustering methods (detailed in supp.), particularly the lack of general-purpose methods, we introduce a Non-Parametric, Plug-and-Play (NP3) algorithm that refines any given cluster partitions using pairwise constraints. Our method first enforces must-link (ML) constraints by merging clusters, followed by a systematic purification process that is used to resolve any resulting cannot-link (CL) violations.

For each impure cluster, we first consider only the data points and constraints within that cluster. Using only these internal constraints, we define a ML group as a set of points that must be clustered together, based on the transitive closure of the ML constraints. We then build a conflict graph where each node represents one of these local ML groups and an edge connects any two groups with a CL constraint between them. The connected components of this conflict graph are our CL groups for the impure cluster.

Since a conflict graph's chromatic number represents the minimum number of partitions needed to satisfy all ML and CL constraints, we can use graph coloring to purify our clusters. We therefore assign cluster labels to the cannot-link group with the highest chromatic number by coloring, and use a Hungarian matching algorithm to assign labels to the remaining cannot-link groups of the impure cluster. This process, which is detailed in the supplementary, results in a final clustering that satisfies all constraints (Fig. 2).

Key Takeaways: Our core idea is to sample **diverse** pairs to address both **under-** and **over-segmentation** errors. We achieve this by first identifying "uncertain regions" (S) that span the entire feature space. Our sampling pool, $\mathcal{U}$, is composed of pairs both *within* ($\mathcal{U}_{us}$) and *across* ($\mathcal{U}_{os}$) these regions, which promotes diversity across the feature space.

- **Over-segmentation** is addressed by querying pairs from $\mathcal{U}_{os}$ to determine if nearby images labeled as different are actually the same.
- **Under-segmentation** is addressed by querying pairs from $\mathcal{U}_{us}$ to determine if neighboring images merged by one method should, in fact, be separate.

This dual-sampling strategy covers most clustering ambiguities while maintaining diverse data coverage.

Algorithm 1: AAS Framework for Training Re-ID Models

Require: Unlabeled images $\mathcal{D}$, feature extractor f, clustering algorithms A, B, budget $\mathcal{B}$

1: Extract features $\mathcal{X} = \{f(x_i) \mid x_i \in \mathcal{D}\}$.

2: **Step 1: Ambiguity-Aware Sampling**

(a) Generate clusters $\mathcal{C}^A = A(\mathcal{X})$ and $\mathcal{C}^B = B(\mathcal{X})$, and apply existing ML/CL constraints.

(b) Identify 'regions of uncertainty' $S = \{S_1, \ldots, S_M\}$ from disagreements between $\mathcal{C}^A$ and $\mathcal{C}^B$.

(c) Construct sampling pools for over-segmentation ($\mathcal{U}_{os}$) and under-segmentation ($\mathcal{U}_{us}$).

(d) Define a marginal probability distribution over the total pool $\mathcal{U} = \mathcal{U}_{os} \cup \mathcal{U}_{us}$ (Eq. 6).

(e) Sample $\mathcal{B}$ pairs for annotation using this distribution.

3: **Step 2: Constrained Pseudo-Label Refinement**

(a) Generate pseudo-labels using a base USL method.

(b) Enforce ML constraints by merging clusters.

(c) For each resulting impure cluster, construct a conflict graph of must-link groups.

(d) Resolve CL violations by graph coloring and Hungarian matching

4: **Step 3: Model training** Retrain the Re-ID model with refined pseudo-labels using base USL method.

4 Experimental Setup

Datasets: We tested our method on a collection of 13 wildlife datasets, all sourced from a subset of WildlifeReID-10k (Adam et al. 2025), while ensuring that none of these datasets were used to train the foundation models. Dataset details are deferred to the supplementary, including evaluations on two human Re-ID tasks using the Market-1501 and Person-X datasets for completeness.

Experimental Protocol: Prior animal Re-ID studies (Čermák et al. 2024; Adam et al. 2025) used the full training set as the gallery and the test set as the query. This protocol enabled *open-set* scenarios, where the query could contain individuals not in the gallery. In (Otarashvili et al. 2024), only the test set was used for generating both the gallery and the query set. Since the open-set setting is more realistic for an Animal Re-ID task, we maintain the test set as query and create the gallery set by choosing exemplary images from a subset of the training set. This strategy enables an open-set evaluation as unseen individuals in the test set are absent in the gallery. We create a held out set of randomly chosen 20% of the individuals in the training set, and up to five exemplars per ID were selected from the remaining 80% of the individuals, using MegaDescriptor based similarity. For a fair comparison, we follow the same protocol for every method. Further details are provided in the supplementary.

Evaluation Metrics: For evaluating animal Re-ID performance in a closed-set scenario, we employed conventional metrics typically used in person Re-ID: mean Average Precision (mAP) and Top-$\{1, 3, 5, 10\}$ accuracy. Additionally, we included mean Inverse Negative Penalty (mINP) for Animal Re-ID. The Top-k accuracy was computed independently for each dataset before being averaged. (Adam et al. 2025) noted that the datasets for animal species might be highly imbalanced, both in terms of number of images as well as individuals. They proposed balancing the perfor-

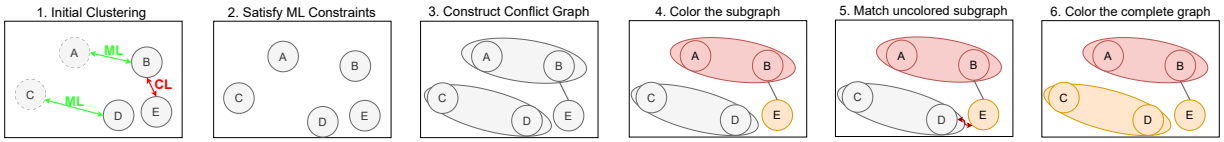


Figure 2: Illustration of the proposed Non-Parametric, Plug-and-Play (NP3) algorithm for refining pseudo labels using pairwise constraints. Given an arbitrary initial clustering, NP3's first step is to satisfy all ML constraints by merging clusters that contain ML pairs. While this addresses ML constraints, it might lead to violations of CL constraints. To purify such impure clusters, our algorithm leverages a combination of graph coloring and Hungarian matching. Intuitively, all ML pairs inside an impure cluster are treated as single conceptual nodes for graph coloring, and any remaining CL pairs in the impure cluster are considered connected via edges. The graph is then colored, ensuring that all CL-constrained pairs are assigned different "colors" and thus separated. For any data points that remain unconstrained by either ML or CL pairs, a Hungarian matching is employed to associate them with their closest existing cluster, a strategy designed to minimize the overall change in the cluster structure.

mance by averaging over all individuals and datasets to provide a general Re-ID performance measure for each individual. Following the same setting, we report Balanced Accuracy on Known Samples (BAKS) to measure the Re-ID performance in closed setting. Similarly, Balanced Accuracy on Unknown Samples (BAUS) uses a fixed threshold on Re-ID scores to determine whether the query is 'known' or 'unknown'. Since a single threshold may not fully reflect overall open-set performance, we instead report the AUC-ROC.

Baselines: We benchmarked the animal Re-ID datasets using two foundation animal Re-ID models (MegaDescriptor (Čermák et al. 2024) and MiewID (Otarashvili et al. 2024)) and multiple USL methods (ICE (Chen, Lagadec, and Bremond 2021), PPLR (Cho et al. 2022), ClusterContrast (Dai et al. 2021), RTMem (Yin et al. 2023), PurificationReID (Lan et al. 2023) and SpCL (Ge et al. 2020)). As mentioned previously, we used SpCL (Ge et al. 2020) as our base USL method for training Re-ID models. For benchmarking existing AL Re-ID methods, due to lack of open-sourced implementation, we re-implemented the most recent and current state-of-the-art AL person Re-ID method, GBAS and LBAS, proposed by (He et al. 2024). We also used a simple baseline that actively samples random pairs of unlabeled images. We also used NIS (Perez et al. 2024), which is not an AL Re-ID method but was proposed as an AL technique designed for animal population estimation. We used the NIS sampling strategy to get oracle feedback for pairwise ML and CL constraints. Similar to all previous USL and AL Re-ID methods, we also use a ResNet-50 backbone to extract the features. We also report baseline (pretrained) and skyline (fine-tuned on 100% data) performances.

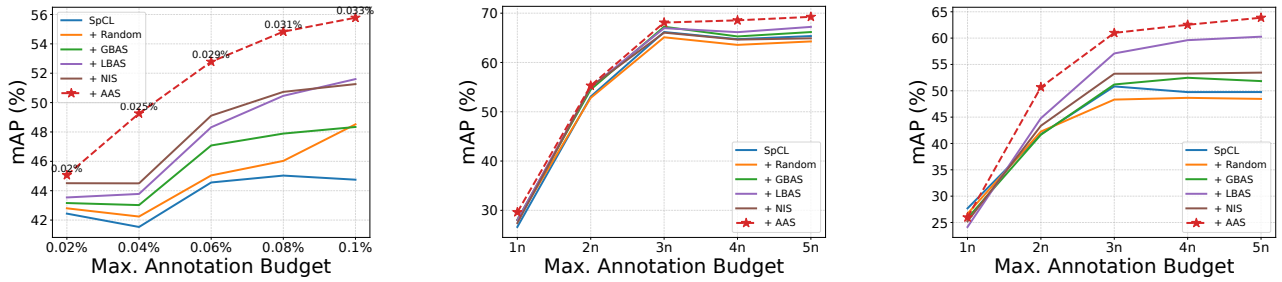
Implementation Details: We use the official code repositories for all foundation, USL and AL Re-ID (only NIS available) methods. Since our AL Re-ID is based on SpCL, we use their official training configuration for training our Re-ID model. This implies that our similarity function $\text{sim}(x_u, x_v)$ is based on the reciprocal K-Nearest Neighbors, as in SpCL. Instead of retraining from scratch after each AL cycle for the specified number of epochs, we train the model once while actively sampling image pairs every few epochs. This reduces the overhead to train the model multiple times for a large number of epochs. This strategy is followed for every AL approach compared. Specifically, we train all our models (USL and AL) for 50 epochs only, and we propose to actively sample the image pairs after every 10 epochs, resulting in a total of 5 AL cycles. Moreover, all AL baselines perform pseudo-label refinement using NP3 for a fair comparison. Unlike person Re-ID, NIS proposed defining the budget using a fraction of total number of all possible pairwise combinations instead of total number of individuals. We also use a similar technique and use a budget 0.02% in each AL cycle. In all our experiments, we fix the hyperparameters $(\epsilon, k_{max}, s_{min})$ to $(0.6, 5, 0.3)$.

5 Results and Analysis

The experimental results demonstrate the superior performance of our proposed Ambiguity-Aware Sampling across both wildlife and person Re-ID benchmarks. We summarize our results on all 13 animal Re-ID datasets in Table 1, where we report the average performance over 4 runs for all AL Re-ID methods. AAS consistently achieves the highest scores across all metrics for closed and open set evaluation, significantly outperforming existing foundational, USL and AL Re-ID methods. We report that by using only 0.033% of all pairwise annotations, we achieve a significant average improvement of 10.49%, 11.19% and 3.99% (mAP) over foundation, USL and AL methods, respectively, while attaining state-of-the-art performance on each dataset. Furthermore, we also show an improvement of 11.09%, 8.2% and 2.06% (AUC ROC) for unknown individuals in an open-world setting. Notably, while all AL methods were allotted a maximum budget of 0.1%, AAS inherently utilized only 0.033% due to its design, further showcasing its efficiency in reducing annotation effort. This is particularly impressive when compared to the 100% budget of the Skyline model, as AAS still closely approaches its performance in key metrics like Top-5 accuracy (85.04% for AAS vs. 85.38% for Skyline), a performance level that other AL methods evidently fail to achieve. In the supplementary, we also show that AAS outperforms LBAS and achieves near skyline performance of 61% mAP with more AL cycles and training iterations, while still sampling only 0.037% of the pairs, even when a much higher budget of 10% is available. We also performed a Wilcoxon signed-rank test on mAP values across all wildlife datasets, obtaining a p-value of at most 1.22e-4 when comparing AAS with all other AL baselines. This

Methods	mAP	mINP	BAKS	AUC ROC	Top-1	Top-3	Top-5	Top-10	Budget
ResNet-50 (Baseline)	38.88%	21.34%	48.62%	60.87%	48.37%	65.21%	71.01%	81.05%	0%
ResNet-50 (Skyline)	64.29%	49.75%	72.04%	80.02%	72.88%	81.06%	85.38%	91.27%	100%
Foundation Methods									
MegaDescriptor	45.65%	26.57%	55.57%	64.12%	56.46%	70.83%	78.14%	85.48%	0%
MiewID	41.16%	21.08%	57.04%	63.72%	56.41%	69.88%	76.63%	85.28%	0%
Unsupervised Learning (USL) Methods									
ICE	48.70%	30.08%	61.17%	70.55%	61.52%	73.93%	80.2%	87.32%	0%
PPLR	51.23%	32.62%	61.99%	71.49%	63.78%	78.19%	81.25%	87.89%	0%
ClusterContrast	45.92%	27.81%	57.97%	69.43%	58.15%	71.83%	77.74%	86.49%	0%
RTMem	48.00%	29.28%	60.57%	69.70%	61.4%	72.23%	77.65%	84.85%	0%
PurificationReID	51.15%	32.63%	62.86%	72.40%	63.16%	76.92%	83.25%	88.45%	0%
Active Learning Methods (based on SpCL)									
SpCL (Base)	44.95%	27.37%	55.38%	67.01%	56.65%	71.04%	76.35%	84.32%	0%
+ Random	46.67%	29.02%	57.28%	68.33%	58.11%	71.96%	78.35%	85.03%	0.1%
+ GBAS	48.61%	30.35%	60.53%	71.50%	60.56%	73.89%	80.06%	87.41%	0.1%
+ LBAS	52.15%	34.06%	63.62%	73.15%	63.99%	76.64%	82.44%	88.18%	0.1%
+ NIS	50.72%	32.73%	61.21%	72.30%	61.78%	75.58%	80.44%	87.41%	0.1%
+ AAS	**56.14%**	**38.17%**	**67.15%**	**75.21%**	**67.71%**	**79.08%**	**85.04%**	**91.18%**	**0.033%**

Table 1: Averaged results for 13 animal Re-ID datasets. Each block reports performance using foundational, USL and AL Re-ID methods, respectively. SpCL is the base method used for training the Re-ID model after each AL cycle.



(a) Wildlife Datasets (b) Market-1501 (c) Person-X

Figure 3: Performance comparison between base SpCL, existing AL Re-ID methods and AAS after every AL cycle. The numbers for AAS in (a) represent the total utilized budget. Since the utilized budget is approximately the same for all AL methods in (b) and (c), we remove them for clarity.

highlights that our improvements are statistically significant.

In Fig 3, we demonstrate that we also outperform all AL-ReID and base SpCL methods after every AL cycle, showcasing that the pairs sampled by AAS are indeed informative. Furthermore, the budget utilized by AAS in each AL cycle reduces as the training progresses. This indicates that the pseudo-labels in the initial phase of training are noisier and therefore require more supervision. We also provide an analysis of the performance for each of the 13 wildlife datasets in the supplementary, with plots on average performance along with the standard deviation.

Our analysis on person Re-ID, presented in the supplementary, reinforces the robustness of AAS. We again observe that we consistently outperform all the evaluated methods across both Market-1501 and Person-X datasets. On Market-1501, we observe an improvement of 1.95% in mAP and 1.05% in Top-1 accuracy over the best performing and state-of-the-art AL strategy LBAS. Similarly, for the Person-X dataset, we observe an improvement of 3.77% and 2.68% in terms of mAP and Top-1 accuracy, respectively, as compared to LBAS. Collectively, these results establish AAS as a highly effective AL strategy that significantly boosts performance while drastically reducing the labeling budget required for both wildlife and person Re-ID applications.

6 Discussion and Conclusion

In this work, we presented a novel AL Re-ID framework, AAS, that targets resolving both *over-* and *under-* segmentation errors by leveraging clustering disagreements. We also introduced NP3, a general-purpose, non-parametric constrained clustering algorithm that seamlessly integrates pairwise constraints into any clustering pipeline. Our experiments demonstrate that AAS outperforms existing foundation, USL and AL Re-ID techniques on a wide range of wildlife and person Re-ID benchmarks. Importantly, AAS achieves this with only 0.033% budget—highlighting its efficiency and scalability. Our work highlights the promise of structured uncertainty modeling in AL and contributes a practical, effective framework for real-world Re-ID systems.

Acknowledgment

This work was supported by ANRF (SERB), Govt. of India, under grant no. CRG/2020/006049. The authors acknowledge the compute infrastructure support from the Infosys Center for Artificial Intelligence at IIIT-Delhi. The authors would like to thank Anjaneya Sharma from IIIT-Delhi for his support in running additional analysis experiments. Lastly, the authors are grateful to the team at Tiger Cell, Wildlife Institute of India, Dehradun, India, for the collaboration that led to this work.

References

Adam, L.; Čermák, V.; Papafitsoros, K.; and Picek, L. 2025. WildlifeReID-10k: Wildlife re-identification dataset with 10k individual animals. arXiv:2406.09211.

Bertulli, C. G.; Rasmussen, M. H.; and Rosso, M. 2016. An assessment of the natural marking patterns used for photo-identification of common minke whales and white-beaked dolphins in Icelandic waters. *Journal of the Marine Biological Association of the United Kingdom*, 96(4): 807–819.

Block, B. A.; Dewar, H.; Blackwell, S. B.; Williams, T.; Prince, E.; Boustany, A. M.; Farwell, C.; Dau, D. J.; and Seitz, A. 2001. *Archival and Pop-up Satellite Tagging of Atlantic Bluefin Tuna*, 65–88. Dordrecht: Springer Netherlands. ISBN 978-94-017-1402-0.

Bolger, D. T.; Morrison, T. A.; Vance, B.; Lee, D.; and Farid, H. 2012. A computer-assisted system for photographic mark–recapture analysis. *Methods in Ecology and Evolution*, 3(5): 813–822.

Carroll, J. M.; Hamm, R. L.; Hagen, J. M.; Davis, C. A.; and Guthery, F. S. 2017. Evaluation of leg banding and attachment of radio-transmitters on ring-necked pheasant chicks. *Wildlife Biology*, 2017(1): 1–6.

Cermak, V.; Picek, L.; Adam, L.; Neumann, L.; and Matas, J. 2025. WildFusion: Individual Animal Identification with Calibrated Similarity Fusion. In Del Bue, A.; Canton, C.; Pont-Tuset, J.; and Tommasi, T., eds., *Computer Vision – ECCV 2024 Workshops*, 18–36. Cham: Springer Nature Switzerland.

Chen, H.; Lagadec, B.; and Bremond, F. 2021. Ice: Inter-instance contrastive encoding for unsupervised person re-identification. In *Proceedings of the IEEE/CVF international conference on computer vision*, 14960–14969.

Cho, Y.; Kim, W. J.; Hong, S.; and Yoon, S.-E. 2022. Part-based pseudo label refinement for unsupervised person re-identification. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 7308–7318.

Cooke, S. J.; Hinch, S. G.; Wikelski, M.; Andrews, R. D.; Kuchel, L. J.; Wolcott, T. G.; and Butler, P. J. 2004. Biotelemetry: a mechanistic approach to ecology. *Trends in Ecology & Evolution*, 19(6): 334–343.

Crall, J. P.; Stewart, C. V.; Berger-Wolf, T. Y.; Rubenstein, D. I.; and Sundaresan, S. R. 2013. HotSpotter — Patterned species instance recognition. In *2013 IEEE Workshop on Applications of Computer Vision (WACV)*, 230–237.

Dai, Z.; Wang, G.; Zhu, S.; Yuan, W.; and Tan, P. 2021. Cluster Contrast for Unsupervised Person Re-Identification. *arXiv preprint arXiv:2103.11568*.

Das, A.; Chakraborty, A.; and Roy-Chowdhury, A. K. 2014. Consistent Re-identification in a Camera Network. In Fleet, D.; Pajdla, T.; Schiele, B.; and Tuytelaars, T., eds., *Computer Vision – ECCV 2014*, 330–345. Cham: Springer International Publishing. ISBN 978-3-319-10605-2.

Das, A.; Panda, R.; and Roy-Chowdhury, A. 2015. Active image pair selection for continuous person re-identification. In *2015 IEEE International Conference on Image Processing (ICIP)*, 4263–4267.

Ester, M.; Kriegel, H.-P.; Sander, J.; and Xu, X. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, KDD'96, 226–231. AAAI Press.

Ge, Y.; Zhu, F.; Chen, D.; Zhao, R.; and Li, H. 2020. Self-paced Contrastive Learning with Hybrid Memory for Domain Adaptive Object Re-ID. In *Advances in Neural Information Processing Systems*.

Guo, S.; Xu, P.; Miao, Q.; Shao, G.; Chapman, C. A.; Chen, X.; He, G.; Fang, D.; Zhang, H.; Sun, Y.; Shi, Z.; and Li, B. 2020. Automatic Identification of Individual Primates with Deep Learning Techniques. *iScience*, 23(8): 101412.

Hammond, P.; Mizroch, S.; and Donovan, G. 1990. Individual recognition of cetaceans: use of photo-identification and other techniques to estimate population parameters. *Reports of the International Whaling Commission, Special Issue*, 12.

Han, P.; Li, Q.; Ma, C.; Xu, S.; Bu, S.; Zhao, Y.; and Li, K. 2021. HMMN: Online metric learning for human re-identification via hard sample mining memory network. *Engineering Applications of Artificial Intelligence*, 106: 104489.

He, T.; Shen, L.; Ding, G.; Zhou, Z.; Xu, T.; Jin, X.; and Huang, Y. 2024. Balanced Active Sampling for Person Re-identification. In *2024 IEEE International Conference on Multimedia and Expo (ICME)*, 1–6.

Hu, B.; Zha, Z.-J.; Liu, J.; Zhu, X.; and Xie, H. 2021. Cluster and Scatter: A Multi-grained Active Semi-supervised Learning Framework for Scalable Person Re-identification. In *Proceedings of the 29th ACM International Conference on Multimedia*, MM '21, 2605–2614. New York, NY, USA: Association for Computing Machinery. ISBN 9781450386517.

Jiao, B.; Liu, L.; Gao, L.; Wu, R.; Lin, G.; WANG, P.; and Zhang, Y. 2023. Toward Re-Identifying Any Animal. In *Thirty-seventh Conference on Neural Information Processing Systems*.

Joshi, V.; and Subramanyam, A. V. 2022. Contextual Active Learning for Person Re- Identification. In *2022 IEEE 5th International Conference on Multimedia Information Processing and Retrieval (MIPR)*, 252–257.

Lan, L.; Teng, X.; Zhang, J.; Zhang, X.; and Tao, D. 2023. Learning to purification for unsupervised person re-identification. *IEEE Transactions on Image Processing*, 32: 3338–3353.

Li, P.; Wu, K.; Zhou, S.; Huang, Q.; and Wang, J. 2023. Pseudo Labels Refinement with Intra-Camera Similarity for Unsupervised Person Re-Identification. In *2023 IEEE International Conference on Image Processing (ICIP)*, 366–370.

Lin, J.; Ren, L.; Lu, J.; Feng, J.; and Zhou, J. 2017. Consistent-Aware Deep Learning for Person Re-identification in a Camera Network. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3396–3405.

Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; and Guo, B. 2021. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*.

Liu, Z.; Wang, J.; Gong, S.; Tao, D.; and Lu, H. 2019. Deep Reinforcement Active Learning for Human-in-the-Loop Person Re-Identification. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 6121–6130.

Lu, W.; Zhao, Y.; Wang, J.; Zheng, Z.; Feng, L.; and Tang, J. 2023. MammalClub: An Annotated Wild Mammal Dataset for Species Recognition, Individual Identification, and Behavior Recognition. *Electronics*, 12(21).

Otarashvili, L.; Subramanian, T.; Holmberg, J.; Levenson, J. J.; and Stewart, C. V. 2024. Multispecies Animal Re-ID Using a Large Community-Curated Dataset. arXiv:2412.05602.

Perez, G.; Sheldon, D.; Van Horn, G.; and Maji, S. 2024. Human-in-the-Loop Visual Re-ID for Population Size Estimation. In *European Conference on Computer Vision (ECCV)*.

Ravela, S.; duyck, j.; finn, c.; Vera, P.; Salas, J.; and hutcheon, a. 2014. Sloop: A pattern retrieval engine for individual animal identification. *Pattern Recognition*, 48.

Roy, S.; Paul, S.; Young, N. E.; and Roy-Chowdhury, A. K. 2018. Exploiting Transitivity for Learning Person Re-identification Models on a Budget. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7064–7072.

Sarfraz, M. S.; Sharma, V.; and Stiefelhagen, R. 2019. Efficient Parameter-free Clustering Using First Neighbor Relations. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 8934–8943.

Schneider, S.; Taylor, G. W.; Linquist, S. S.; and Kremer, S. C. 2018. Past, Present, and Future Approaches Using Computer Vision for Animal Re-Identification from Camera Trap Data. arXiv:1811.07749.

Shukla, A.; Anderson, C.; Cheema, G. S.; Gao, P.; Onda, S.; Anshumaan, D.; Anand, S.; and Farrell, R. 2019a. A Hybrid Approach to Tiger Re-Identification. *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, 294–301.

Shukla, A.; Cheema, G. S.; Anand, S.; Qureshi, Q.; and Jhala, Y. 2019b. Primate Face Identification in the Wild. arXiv:1907.02642.

Tuia, D.; Kellenberger, B.; Beery, S.; Costelloe, B. R.; Zuffi, S.; Risse, B.; Mathis, A.; Mathis, M. W.; van Langevelde, F.; Burghardt, T.; Kays, R.; Klinck, H.; Wikelski, M.; Couzin, I. D.; van Horn, G.; Crofoot, M. C.; Stewart, C. V.; and Berger-Wolf, T. 2022. Perspectives in machine learning for wildlife conservation. *Nature Communications*, 13(1): 792.

Čermák, V.; Picek, L.; Adam, L.; and Papafitsoros, K. 2024. WildlifeDatasets: An Open-Source Toolkit for Animal Re-Identification. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 5953–5963.

Walker, K. A.; Trites, A. W.; Haulena, M.; and Weary, D. M. 2012. A review of the effects of different marking and tagging techniques on marine mammals. *Wildlife Research*, 39(1): 15–30.

Wang, H.; Gong, S.; Zhu, X.; and Xiang, T. 2018. Human-In-The-Loop Person Re-Identification. arXiv:1612.01345.

Wang, M.; Lai, B.; Huang, J.; Gong, X.; and Hua, X.-S. 2021. Camera-aware Proxies for Unsupervised Person Re-Identification. arXiv:2012.10674.

Wolcott, T. G. 1995. New options in physiological and behavioural ecology through multichannel telemetry. *Journal of Experimental Marine Biology and Ecology*, 193(1): 257–275. Behavioural Ecology of Decapod Crustaceans: An Experimental Approach.

Wu, Y.; Huang, T.; Yao, H.; Zhang, C.; Shao, Y.; Han, C.; Gao, C.; and Sang, N. 2021. Multi-Centroid Representation Network for Domain Adaptive Person Re-ID. arXiv:2112.11689.

Yin, J.; Zhang, X.; Ma, Z.; Guo, J.; and Liu, Y. 2023. A Real-Time Memory Updating Strategy for Unsupervised Person Re-Identification. *IEEE Transactions on Image Processing*, 32: 2309–2321.

Zhang, X.; Ge, Y.; Qiao, Y.; and Li, H. 2021. Refining Pseudo Labels with Clustering Consensus over Generations for Unsupervised Object Re-identification. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 3435–3444.

Zheng, Y.; Tang, S.; Teng, G.; Ge, Y.; Liu, K.; Qin, J.; Qi, D.; and Chen, D. 2021. Online Pseudo Label Generation by Hierarchical Cluster Dynamics for Adaptive Person Re-identification. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 8351–8361.

Zou, C.; Chen, Z.; Cui, Z.; Liu, Y.; and Zhang, C. 2023. Discrepant and Multi-instance Proxies for Unsupervised Person Re-identification. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 11024–11034.

Reproducibility Checklist

1. General Paper Structure

1.1. Includes a conceptual outline and/or pseudocode description of AI methods introduced (yes/partial/no/NA) yes

1.2. Clearly delineates statements that are opinions, hypothesis, and speculation from objective facts and results (yes/no) yes

1.3. Provides well-marked pedagogical references for less-familiar readers to gain background necessary to replicate the paper (yes/no) yes

2. Theoretical Contributions

2.1. Does this paper make theoretical contributions? (yes/no) no

3. Dataset Usage

3.1. Does this paper rely on one or more datasets? (yes/no) yes

If yes, please address the following points:

3.2. A motivation is given for why the experiments are conducted on the selected datasets (yes/partial/no/NA) yes

3.3. All novel datasets introduced in this paper are included in a data appendix (yes/partial/no/NA) NA

3.4. All novel datasets introduced in this paper will be made publicly available upon publication of the paper with a license that allows free usage for research purposes (yes/partial/no/NA) NA

3.5. All datasets drawn from the existing literature (potentially including authors' own previously published work) are accompanied by appropriate citations (yes/no/NA) yes

3.6. All datasets drawn from the existing literature (potentially including authors' own previously published work) are publicly available (yes/partial/no/NA) yes

3.7. All datasets that are not publicly available are described in detail, with explanation why publicly available alternatives are not scientifically satisficing (yes/partial/no/NA) NA

4. Computational Experiments

4.1. Does this paper include computational experiments? (yes/no) yes

If yes, please address the following points:

4.2. This paper states the number and range of values tried per (hyper-) parameter during development of the paper, along with the criterion used for selecting the final parameter setting (yes/partial/no/NA) yes

4.3. Any code required for pre-processing data is included in the appendix (yes/partial/no) no

4.4. All source code required for conducting and analyzing the experiments is included in a code appendix (yes/partial/no) no

4.5. All source code required for conducting and analyzing the experiments will be made publicly available upon publication of the paper with a license that allows free usage for research purposes (yes/partial/no) yes

4.6. All source code implementing new methods have comments detailing the implementation, with references to the paper where each step comes from (yes/partial/no) yes

4.7. If an algorithm depends on randomness, then the method used for setting seeds is described in a way sufficient to allow replication of results (yes/partial/no/NA) yes

4.8. This paper specifies the computing infrastructure used for running experiments (hardware and software), including GPU/CPU models; amount of memory; operating system; names and versions of relevant software libraries and frameworks (yes/partial/no) partial

4.9. This paper formally describes evaluation metrics used and explains the motivation for choosing these metrics (yes/partial/no) yes

4.10. This paper states the number of algorithm runs used to compute each reported result (yes/no) yes

4.11. Analysis of experiments goes beyond single-dimensional summaries of performance (e.g., average; median) to include measures of variation, confidence, or other distributional information (yes/no) yes

4.12. The significance of any improvement or decrease in performance is judged using appropriate statistical tests (e.g., Wilcoxon signed-rank) (yes/partial/no) yes

4.13. This paper lists all final (hyper-)parameters used for each model/algorithm in the paper's experiments (yes/partial/no/NA) yes