A Survey on Animal Re-Identification: Advances, Datasets, and Challenges (2020–2025)

Author
UniVR
Email: test.email

***Abstract*—Animal re-identification (Re-ID) has emerged as a critical task in computer vision, enabling the recognition of individual animals across images or videos [1]. It plays a vital role in biodiversity monitoring, behavioral ecology, and conservation management. Building upon advances in person Re-ID, recent research has proposed novel architectures, learning strategies, and benchmark datasets tailored to the challenges of animal identification — including inter-species variation, limited labeled data, and dynamic environments.**

This survey provides a comprehensive review of animal Re-ID research from 2020 to 2025. We cover both *closed-world* approaches, where all individuals are known at training time, and *open-world* settings, which require recognizing unseen identities in the wild. The survey analyzes state-of-the-art techniques for feature representation learning (e.g., CNNs, Vision Transformers, and multimodal fusion), similarity learning (e.g., metric losses, few-shot learning, and unsupervised methods), and generalization strategies across species and domains.

In addition, we examine major datasets and benchmarks — including WildlifeReID-10k, SeaTurtleID2022, and OpenAnimals — and summarize evaluation metrics and protocols such as CMC, mAP, and open-set verification. Finally, we highlight open challenges and future directions, including species-agnostic modeling, metadata integration, continual learning, and ethical considerations in ecological deployment.

By systematically organizing the literature and identifying key trends, this survey aims to support future research in building robust, scalable, and ethically responsible animal Re-ID systems.

I. Introduction

Animal re-identification (Re-ID) is the task of recognizing individual animals across multiple images or video sequences captured under varying conditions. Unlike object classification, which assigns an image to a predefined category (e.g., species), animal Re-ID aims to differentiate between instances within the same category — identifying a specific animal regardless of pose, lighting, background, or camera viewpoint.

This capability is vital for numerous ecological and conservation applications, such as behavioral analysis, population monitoring, and anti-poaching surveillance. Manual identification by experts is time-consuming, error-prone, and infeasible at scale, making automated Re-ID an essential tool for modern wildlife research.

A. Motivation and Background

Recent years have seen growing interest in adapting person Re-ID methodologies to animals. Early works focused on species-specific pattern matching, such as stripe and spot analysis in zebras and leopards [2], [3], while more recent approaches have leveraged deep learning to extract robust, individual-specific representations [4], [5], [6].

Several factors have contributed to this growth:

- The proliferation of camera traps, drones, and underwater video systems, allowing large-scale, non-invasive data collection [7], [8]
- Advances in computer vision — particularly CNNs [9], transformers [10], and vision-language models [11]
- New datasets spanning multiple species and environments [12], [13]

Yet, animal Re-ID introduces challenges that go beyond human Re-ID. These include:

- Drastic appearance variation within species due to pose, lighting, and seasonal change
- Scarcity of annotated data for many species, especially in open environments
- Difficulty in defining inter-class boundaries among similar-looking individuals

B. Survey Scope and Contributions

This survey provides a comprehensive overview of recent progress in animal Re-ID between 2020 and 2025. We analyze the field from both methodological and ecological perspectives and highlight emerging techniques across the following dimensions:

- **Closed-world vs. Open-world Re-ID:** We distinguish methods designed for known identity sets (e.g., farm or zoo animals) from those operating in dynamic, open environments (e.g., wild populations) [1], [14], [13].
- **Feature Learning:** We review CNN-based methods [4], [6], attention mechanisms [15], and transformer architectures [16], [17] used for individual-level discrimination.
- **Similarity Learning:** We cover metric learning techniques such as contrastive loss [18], triplet loss [19], few-shot and meta-learning [20], and hybrid objectives [21].
- **Benchmarks and Evaluation:** We summarize datasets such as WildlifeReID-10k [12], SeaTurtleID2022 [7], SealID [8], and OpenAnimals [13], and discuss evaluation protocols including open-set recognition [22], clustering [23], and multimodal assessment [24].
- **Challenges and Future Directions:** We identify key obstacles — such as species generalization, lack of labels, and benchmark fragmentation — and suggest promising research paths, including continual learning and ethical deployment [25], [20], [11].

To the best of our knowledge, this is the first survey to comprehensively synthesize animal Re-ID literature across both closed- and open-world paradigms, with an emphasis on methodological rigor and real-world applicability.

C. Structure of the Survey

The remainder of this paper is organized as follows: Section II presents closed-world animal Re-ID systems. Section **??** covers open-world and generalizable Re-ID methods. Section **??** focuses on feature representation learning. Section **??** discusses similarity learning and loss functions. Section **??** reviews datasets and benchmarks. Section **??** outlines evaluation metrics and protocols. Section **??** explores open challenges and research directions. Finally, Section **??** concludes the survey and presents an outlook.

II. Closed-World Animal Re-Identification

Closed-world animal re-identification (Re-ID) refers to the task of distinguishing between a fixed and predefined set of known animal identities. Unlike open-world scenarios, the system is not expected to generalize to novel individuals but rather to recognize and retrieve from a closed gallery of labeled identities. This setting is commonly adopted in zoo environments, farm monitoring, or controlled conservation efforts, where individuals are already labeled and the system can be trained on all target identities.

A. Problem Definition and Motivation

In a closed-world scenario, all animal identities present during inference have also been observed during training. The challenge is to robustly distinguish between known individuals despite intra-class variation due to pose, lighting, background, and occlusion. This problem is similar in formulation to traditional classification, yet requires learning features that are discriminative across fine-grained individual differences rather than categorical labels.

B. Architectures and Techniques

The majority of closed-world animal Re-ID systems adopt architectures derived from person Re-ID or fine-grained classification pipelines, leveraging Convolutional Neural Networks (CNNs) or Vision Transformers (ViTs). For instance, classification-based methods typically model the problem as multi-class classification with softmax cross-entropy loss, treating each individual as a separate class [9]. However, metric learning remains dominant, particularly via Siamese or triplet architectures [18], [19].

Several works adopt or extend Siamese networks for animal Re-ID in closed settings. For example, Verma et al. [6] propose a Siamese-based knowledge distillation framework for wildlife identification, showing competitive performance with limited data. Similarly, Li et al. [2] use multi-image feature aggregation combined with geometric similarity to boost recognition accuracy in patterned animals like zebras or tigers.

Classification-based approaches are still employed effectively in controlled datasets. For instance, the Adaptive Feature Fusion Network (AFFN) by Feng et al. [5] combines multiple visual streams with a final softmax head trained with cross-entropy. Other works, such as Chang et al. [26], utilize multi-scale fusion techniques to enhance discriminative features across body parts, particularly useful in species like the Amur tiger.

C. Feature Representation in Closed Settings

Deep CNNs like ResNet [9], Inception [**?**], and hybrid models remain foundational. Transformer-based embeddings are also increasingly used. Kim et al. [27] utilize ViTs for lion Re-ID in a closed-world setting, achieving strong performance on benchmark datasets. Bhattacharya et al. [4] combine local and global representations to capture inter-individual differences.

Some works apply part-based or landmark-guided learning. Yu et al. [28] propose landmark-guided embeddings to enhance part-level distinctions. This technique is particularly beneficial in structured species like zebras and giraffes, where body landmarks correspond to distinct pattern signatures.

D. Benchmark Datasets

Closed-world models are typically evaluated on datasets where all test identities are seen during training. Notable datasets include:

- **ZebraID** [3]: Contains labeled images of individual zebras in the wild.
- **YakRe-ID-103** [29]: Designed for yak identification under farm conditions with consistent lighting and camera angles.
- **SealID** [8]: Focused on the Saimaa ringed seal in a conservation context.
- **SeaTurtleID2022** [7]: Provides long-term identity labels across various lighting and occlusion conditions.

These datasets often provide sufficient per-individual samples to train effective discriminative classifiers or metric embedding functions.

E. Evaluation and Limitations

Evaluation in the closed-world setting commonly uses Cumulative Matching Characteristic (CMC) and mean Average Precision (mAP) [17]. While closed-world setups are easier to train and evaluate, they suffer from key limitations:

- **Overfitting to Known Individuals:** Models may memorize training identities without learning generalizable features.
- **Poor Scalability:** Adding new individuals requires retraining or architectural adaptation.
- **Limited Ecological Validity:** Real-world deployments often encounter unknown or unlabeled individuals.

Despite these limitations, closed-world Re-ID remains crucial in scenarios with fixed populations and well-labeled individuals, such as farm management or captive breeding programs [24], [30].

III. Open-World Animal Re-Identification

Open-world animal re-identification (Re-ID) extends the challenge of identifying individual animals to unconstrained environments where novel identities may appear during deployment. Unlike the closed-world setting, the system must be capable of rejecting unknown individuals or incorporating them incrementally without full retraining. This setting reflects real-world ecological scenarios, where wild populations are dynamic and unlabeled.

A. Problem Setting and Motivation

In the open-world setting, test samples may include both known and unseen individuals. The task thus requires the model to (i) identify individuals already present in the gallery, and (ii) detect unknown identities and possibly assign them new labels. Such functionality is critical for wildlife monitoring, where new individuals (e.g., newborns, migrants, reappearing animals) are routinely encountered [14], [17], [13].

Compared to closed-world Re-ID, open-world scenarios present additional challenges, such as:

- Recognition under limited prior knowledge
- Handling long-tailed or imbalanced identity distributions
- Incrementally learning from new individuals
- Avoiding false positives among known classes

B. Open-Set Recognition and Outlier Detection

A key capability of open-world Re-ID models is to detect when a query does not belong to any known identity. Several works employ outlier detection or open-set recognition techniques for this purpose.

Qian et al. [7] explore temporal and spatial patterns in SeaTurtleID2022 to separate new arrivals from previously observed individuals. Frühner and Tapken [22] adapt person Re-ID techniques to multispecies animal domains and emphasize the importance of outlier-aware models in field deployment.

Liu et al. [31] propose a locally aware transformer with cross-attention to generalize across diverse animal appearances, useful for recognizing out-of-distribution samples. Unsupervised domain adaptation and novelty-aware embeddings are increasingly integrated into such systems.

C. Few-Shot and Meta Learning Approaches

Because many open-world deployments have limited labeled examples for new individuals, few-shot and meta-learning techniques are gaining traction. Choudhury et al. [20] introduce a meta-feature adapter that integrates environmental metadata to improve few-shot adaptation. Similarly, Alahi et al. [1] propose a few-shot learning framework leveraging metric learning and hard sample mining to generalize to unseen animal identities.

Prototypical networks and model-agnostic meta-learning (MAML) offer effective solutions by enabling fast adaptation to new classes. These are especially valuable for long-term monitoring where retraining is impractical.

D. Unsupervised and Self-Supervised Learning

Open-world Re-ID also requires robust representations that generalize without overfitting to labeled individuals. Several works pursue unsupervised or self-supervised approaches.

Andersson et al. [25] propose a fully unsupervised pipeline that unwraps pelage patterns to learn representations invariant to pose and viewpoint. Zhao et al. [21] present IndivAID, which disentangles appearance representations using adversarial losses to boost unsupervised identification.

Cluster-based pseudo-labeling [23], domain-adversarial learning, and contrastive pretraining (e.g., MoCo, SimCLR) are also increasingly adopted in recent open-world systems.

E. Cross-Species and Multispecies Generalization

A hallmark of the open-world paradigm is cross-species generalization — the ability to re-identify individuals across different species or domains. Xie et al. [13] curate a large-scale multi-species dataset to facilitate such generalization. Wang et al. [30] propose an unsupervised part-based alignment method that improves robustness under inter-species variation.

Jiao et al. [14] tackle the open challenge of re-identifying any animal in the wild by using a generalizable transformer-based model. Their method leverages multi-scale attention and hierarchical fusion to extract discriminative features across species boundaries.

F. Video-Based and Temporal Open-World Re-ID

Video data enables tracking of new individuals over time. Hansen et al. [32] and Jensen et al. [23] propose track-level clustering for unsupervised discovery and incremental labeling. These temporal signals help to group images of unseen animals and assign new identities with confidence.

Verma et al. [6] use temporal consistency for pseudo-label refinement in knowledge-distilled Siamese networks, allowing for more accurate few-shot recognition in the wild.

G. Challenges and Opportunities

Open-world Re-ID systems offer ecological scalability but remain limited by:

- Lack of benchmark protocols simulating realistic open-world settings
- Poor cross-dataset generalization
- Ambiguity in identity boundaries (e.g., visually similar individuals)
- Difficulty in evaluation of novel identity detection

Nonetheless, they present exciting opportunities for biodiversity monitoring, long-term conservation, and adaptive population studies. Future directions include vision-language models for semantic similarity [11], joint appearance-context modeling, and continual learning systems that adapt over time.

IV. Feature Representation Learning

Feature representation learning forms the backbone of animal re-identification (Re-ID) systems. Unlike categorical classification, animal Re-ID requires learning fine-grained, individual-specific features that are invariant to pose, illumination, occlusion, and background variations. Robust representations enable the system to match individuals across space and time, regardless of viewpoint or appearance changes.

A. CNN-Based Feature Extractors

Convolutional Neural Networks (CNNs) have long been the foundation for visual representation learning. Architectures like ResNet [9] and Inception have been widely adopted in animal Re-ID, often serving as backbones for metric learning frameworks.

Feng et al. [5] propose an Adaptive Feature Fusion Network (AFFN) for Amur tiger Re-ID, combining multiple CNN branches to extract global and local features. Similarly, Verma et al. [6] employ Siamese networks with ResNet encoders for knowledge-distilled embeddings under limited data. Bhattacharya et al. [4] aggregate CNN features across image pairs and integrate geometric alignment for improved robustness.

Fine-tuning these networks on animal-specific datasets such as WildlifeReID-10k [12] or ZebraID [3] significantly improves their discriminative ability for individual-level Re-ID.

B. Part-Based and Landmark-Guided Representations

To enhance robustness under occlusion and pose variations, part-based models and landmark-guided embeddings have been adopted.

Yu et al. [28] introduce a landmark-guided feature extractor, enabling part-aware learning in species with structured appearances (e.g., zebras, leopards). Chang et al. [26] propose a serial multi-scale feature fusion mechanism that captures hierarchical patterns across body parts in tigers. Wang et al. [30] extend this idea using unsupervised part alignment across views, enhancing inter-instance discrimination.

Attention mechanisms, including Squeeze-and-Excitation (SE) [33] and CBAM [15], are often integrated to selectively focus on the most informative spatial or channel-wise features. These modules are lightweight yet effective additions to CNN-based architectures.

C. Pattern-Specific Descriptors

Many animals exhibit individual-specific patterns such as stripes, spots, or scars. Methods leveraging these features often outperform holistic models in species like zebras, leopards, and seals.

Andersson et al. [25] propose pelage pattern unwrapping for pose-invariant pattern extraction. Their unsupervised method flattens texture maps of animal coats, enabling the model to isolate identity cues while ignoring background or pose artifacts.

Li et al. [2] and Singh et al. [34] also incorporate pattern similarity and geometric constraints, enabling robust matching under deformations and posture variation. These models outperform global embeddings in structured species where fine pattern features are critical.

D. Vision Transformers and Hybrid Architectures

Transformer-based models have recently emerged as powerful alternatives to CNNs, offering the ability to model long-range dependencies via self-attention.

Dosovitskiy et al.'s ViT [10] serves as the foundation for several works. Kim et al. [27] apply ViTs to lion Re-ID, showing improved generalization over CNNs in cluttered environments. Li et al. [17] introduce an Adaptive High-Frequency Transformer that fuses low- and high-frequency cues, enabling better separation of appearance noise from identity signals.

Xu et al. [16] propose a hierarchical frequency fusion transformer with spatial alignment, outperforming standard ViTs on datasets like SealID [8]. These models benefit from hierarchical tokenization and local-global context fusion, particularly in open environments with appearance variations.

E. Cross-Modal and Auxiliary Feature Learning

To improve generalization in low-resource settings, several works integrate auxiliary or contextual cues alongside visual features.

Choudhury et al. [20] introduce a meta-feature adapter that fuses environmental metadata (e.g., GPS, weather, timestamp) with image features. This context-aware embedding improves few-shot generalization in dynamic environments. Rao et al. [24] provide tools to incorporate such metadata in dataset preprocessing pipelines.

Zhao et al. [21] disentangle appearance from background using adversarial learning, enabling feature representations to focus on identity-specific cues. This is particularly useful when training data contains environmental bias or unbalanced background contexts.

F. Pretraining and Transfer from Person Re-ID

Transfer learning from person Re-ID remains an effective strategy. Pretrained models like FaceNet [18], SphereFace [35], and CosFace [36] provide strong initialization for animal datasets, especially when labeled data is limited.

Frühner and Tapken [22] demonstrate that person Re-ID models can be adapted to multi-species animal scenarios with minor architectural modifications and fine-tuning. Models such as CLIP [11], originally designed for vision-language tasks, are also showing promise in animal Re-ID through cross-modal supervision.

G. Discussion

Overall, animal Re-ID benefits from diverse feature learning paradigms — from handcrafted part alignment to end-to-end attention-based models. CNNs remain effective baselines, especially when augmented with part cues, pattern extractors, or metadata. However, transformers and multimodal embeddings offer new avenues for scalable and generalizable feature extraction in both closed- and open-world settings.

V. Similarity Learning and Loss Functions

In animal re-identification (Re-ID), similarity learning refers to the process of mapping visual inputs into a feature space where samples of the same identity are close, and those of different identities are far apart. Unlike closed-set classification, which focuses on class prediction, similarity learning is essential for generalization to new individuals, few-shot settings, and open-world scenarios.

A. Siamese and Triplet Networks

Siamese and triplet networks are two of the most widely adopted architectures for metric learning. They learn an embedding space using pairwise or triplet image comparisons.

Siamese Networks utilize twin subnetworks sharing weights to compare pairs of images. The network is trained with contrastive loss [18], minimizing the distance between positive pairs (same identity) and maximizing it for negative pairs (different identities). Verma et al. [6] use Siamese architectures combined with knowledge distillation for low-data settings, showing robustness in farm-based animal identification.

Triplet Networks take an anchor, positive, and negative image as input, optimizing the triplet loss to ensure that the positive sample is closer to the anchor than the negative by a margin [19]. This architecture is employed in several animal Re-ID pipelines, including the pattern-aware model by Singh et al. [34], which uses hard triplet mining on posture-diverse images.

B. Contrastive, Cross-Entropy, and Hybrid Losses

Contrastive Loss encourages a margin between positive and negative pairs, making it suitable for Siamese networks [18]. **Triplet Loss** generalizes this to three-sample settings and is widely adopted in open-world applications [19].

Cross-Entropy Loss is used in classification-based approaches where each identity is treated as a separate class. This loss is effective in closed-world scenarios, such as in the CNN-based AFFN model by Feng et al. [5] and the multi-scale fusion model by Chang et al. [26].

Recent works combine multiple loss functions to balance classification accuracy and embedding discriminability. For example:

$$\mathcal{L}_{\text{hybrid}} = \lambda_1 \mathcal{L}_{\text{triplet}} + \lambda_2 \mathcal{L}_{\text{contrastive}} + \lambda_3 \mathcal{L}_{\text{CE}}$$

where λ_i are weighting coefficients for each component.

Such hybrid formulations are explored in Zhao et al.'s IndivAID framework [21], where disentangled representations are learned via joint classification and embedding objectives.

C. Multi-Image Aggregation

Real-world animal datasets often contain multiple images per identity, captured under different poses, lighting, or occlusion conditions. Aggregating features across multiple views can improve embedding robustness.

Li et al. [2] propose a geometric similarity-enhanced multi-image matching method that aligns pose-variant images before aggregation. Bhattacharya et al. [4] fuse visual and spatial features across images to better distinguish individuals with repetitive patterns (e.g., zebras, tigers).

Feature fusion methods include:

- Averaging or weighted pooling of embeddings
- Temporal or attention-based alignment (especially in video)
- Confidence-based feature selection

These strategies reduce noise from individual frames and leverage complementary views for improved performance.

D. Few-Shot and Meta Learning

Few-shot learning aims to generalize to new individuals with very limited samples. It is especially useful in ecological studies where manual annotation is costly.

Alahi et al. [1] introduce a few-shot learning framework based on contrastive learning, optimized with hard pair mining for better generalization. Prototypical networks are also used, where each class prototype is computed as the mean of support embeddings, and classification is done via distance comparison.

Choudhury et al. [20] enhance few-shot adaptation using environmental metadata via a meta-feature adapter. Their model shows improved performance in low-shot learning for rare or newly observed individuals.

E. Unsupervised and Self-Supervised Approaches

Given the scarcity of labeled data, unsupervised and self-supervised learning methods are gaining popularity. Andersson et al. [25] employ unsupervised learning of pelage patterns using pattern unwrapping and clustering. Jensen et al. [23] cluster video tracks to pseudo-label new individuals without manual annotations.

Contrastive self-supervised methods like SimCLR and MoCo can be adapted to animal Re-ID by generating positive pairs via augmentations and pulling them together in the feature space. Zhao et al. [21] further explore adversarial learning to disentangle identity features from background noise.

F. Transfer Learning from Person Re-ID

Due to the maturity of person Re-ID, transferring pretrained embedding models is a common strategy. Frühner and Tapken [22] demonstrate that person Re-ID models such as MGN and PCB can be effectively adapted to multi-species animal Re-ID through fine-tuning.

Pretrained models like FaceNet [18], SphereFace [35], and CosFace [36] provide strong initializations, especially when animal datasets are small. Transfer learning helps accelerate convergence and improve generalization across species.

G. Discussion

Similarity learning enables animal Re-ID systems to scale beyond closed-set classification by learning generalizable embeddings. Techniques such as triplet loss, few-shot prototyping, and unsupervised clustering provide flexible training mechanisms under data-scarce and open-world conditions.

Future directions include curriculum-based metric learning, continual embedding refinement, and integration with vision-language models for semantic-aware similarity spaces [11].

VI. Datasets and Benchmarks

The effectiveness of animal re-identification (Re-ID) models heavily depends on the availability and quality of datasets used for training and evaluation. Unlike human Re-ID, animal Re-ID must deal with a much broader diversity of species, less standardized imaging conditions, and significantly fewer annotated identities. This section reviews key datasets, their properties, and the current benchmarking landscape.

A. Dataset Characteristics and Challenges

Animal Re-ID datasets vary widely in scale, species diversity, and collection context. Compared to person Re-ID benchmarks, animal datasets are typically smaller, less curated, and more prone to class imbalance. Common challenges include:

- **Interspecies Diversity:** Different species require unique feature representations due to differences in body structure, pattern, and color.
- **Environmental Variability:** Datasets include images captured in diverse lighting, weather, and habitat conditions, often with occlusion and motion blur.
- **Limited Annotations:** Collecting per-individual identity labels in the wild is difficult and expensive, especially for rare or elusive animals.
- **Temporal Appearance Change:** Many animals exhibit seasonal coat changes, growth, or aging, which introduces intra-class variation over time.

Several datasets attempt to address these challenges by including long-term tracking, multiple viewpoints, or multi-modal metadata.

B. Notable Datasets

We summarize here the most influential animal Re-ID datasets from 2020–2025:

WildlifeReID-10k [12] is a large-scale dataset containing over 10,000 individual animals across multiple species, including giraffes, elephants, and zebras. It provides a valuable benchmark for cross-species Re-ID and generalization.

SeaTurtleID2022 [7] focuses on long-term monitoring of sea turtles. It includes images from multiple years and environments, enabling temporal Re-ID evaluation.

ZebraID [3] is a pattern-heavy dataset used primarily for benchmarking methods that leverage stripe-based descriptors. It provides a good testbed for unsupervised and few-shot learning techniques.

YakRe-ID-103 [29] features horn-based identity cues in controlled farm environments, useful for testing models under consistent imaging conditions.

SealID [8] includes Saimaa ringed seals and supports cross-season re-identification. Its focus on one endangered species makes it suitable for conservation-driven evaluations.

OpenAnimals [13] is a community-curated dataset designed for large-scale multispecies Re-ID. It contains diverse environments and varying image quality, encouraging research into generalizable and robust models.

IndivAID [21] emphasizes fine-grained appearance-based recognition. It supports evaluation of disentangled representations by including animals with subtle visual differences.

Lion Re-ID (Kim et al.) [27] and **Amur Tiger datasets** [5], [26] are species-specific benchmarks focusing on regional conservation efforts, offering high-quality annotations and species-specific challenges.

C. Multimodal and Contextual Datasets

Rao et al. [24] introduce **WildlifeDatasets**, a toolkit supporting metadata-enhanced datasets. Choudhury et al. [20] leverage GPS, weather, and timestamp information as additional input modalities, enabling evaluation of context-aware models.

Multimodal datasets allow researchers to test models that fuse visual and non-visual information — critical for open-world and few-shot scenarios.

D. Benchmarking Metrics and Evaluation Protocols

Animal Re-ID tasks are typically evaluated using metrics borrowed from person Re-ID:

- **Cumulative Matching Characteristics (CMC):** Measures top-k retrieval accuracy.
- **mean Average Precision (mAP):** Aggregates ranking precision across queries.
- **Top-k Accuracy:** Simpler metric indicating whether the ground-truth identity appears in the top k predictions.
- **AUC and NMI:** Used in open-world and clustering-based evaluation settings [23], [32].

Evaluation protocols often fall into two categories:

- **Closed-Set Evaluation:** All test identities are known during training. This protocol is easier but less realistic for wild environments.
- **Open-Set Evaluation:** Includes unseen individuals during testing. This better simulates wildlife monitoring and requires novelty detection.

Video-based evaluations [23], [32] use track-level identity consistency, leveraging temporal information for robust matching.

E. Challenges in Benchmark Standardization

Unlike human Re-ID, there is no widely accepted benchmark suite for animal Re-ID. Each dataset often comes with its own evaluation splits and protocols. This hinders reproducibility and cross-paper comparison.

- Few datasets include standardized train/test splits.
- Cross-dataset benchmarking is rarely reported.
- Multimodal evaluation pipelines are underdeveloped.

Frühner and Tapken [22] argue for a unified framework to assess species-agnostic and domain-transfer performance across multiple datasets.

F. Future Directions

To advance benchmarking in animal Re-ID, we propose the following:

- **Cross-Species Benchmarks:** New datasets should span multiple species and environments to evaluate generalization.
- **Longitudinal Tracking:** Datasets must include temporal variation for robust Re-ID under aging or seasonal change.
- **Unified Evaluation Protocols:** Standard splits and evaluation scripts will ensure reproducibility.
- **Metadata-Rich Datasets:** Integrating GPS, timestamp, and behavior data will encourage multimodal model design.

Efforts like OpenAnimals [13] and WildlifeDatasets [24] mark an important step in this direction. However, the field still lacks an equivalent to Market-1501 or MSMT17 from the human Re-ID community.

VII. Evaluation Metrics and Protocols

Robust and meaningful evaluation protocols are essential for measuring the performance of animal re-identification (Re-ID) models. While many metrics and methodologies are inherited from person Re-ID, the diversity of species, image modalities, and use cases in animal Re-ID introduces unique challenges. This section outlines the standard evaluation metrics, common protocols, and current limitations.

A. Core Evaluation Metrics

Cumulative Matching Characteristic (CMC): The CMC curve measures the probability that a correct identity appears in the top-k matches returned by the system. It is most appropriate for single-shot closed-set evaluation and is widely reported in animal Re-ID studies [4], [2], [7].

$$\text{CMC@k} = \frac{1}{N}\sum_{i=1}^{N} I(\text{rank}_i \leq k)$$

where $I(\cdot)$ is the indicator function, and rank_i is the rank of the correct match for the i-th query.

Mean Average Precision (mAP): mAP evaluates the precision of retrieval across all recall levels. It is particularly useful when multiple ground truth matches are present for a single query. It is now a standard metric for multi-shot and video-based Re-ID [23], [31].

$$\text{mAP} = \frac{1}{Q}\sum_{q=1}^{Q}\frac{1}{R_q}\sum_{r=1}^{R_q} P(q,r)$$

where Q is the number of queries and R_q is the number of relevant instances for query q.

Top-k Accuracy: This metric indicates whether the correct match appears within the top-k predictions. It is simpler than mAP but still informative, especially in few-shot or metadata-aware Re-ID settings [20].

Area Under the ROC Curve (AUC): AUC is used in open-set Re-ID to evaluate the trade-off between true positive and false positive rates when distinguishing known vs. unknown identities [22], [13].

Normalized Mutual Information (NMI): NMI is used to evaluate clustering quality in unsupervised and self-supervised Re-ID systems. It measures the alignment between predicted clusters and true identities [25], [21].

B. Evaluation Protocols

Closed-Set Evaluation: In this setting, all identities in the test set are also seen during training. This protocol is common in zoo or farm datasets like YakRe-ID-103 [29] and SeaTurtleID2022 [7]. It is often combined with softmax classification or metric learning.

Open-Set Evaluation: Here, the test set includes unknown individuals not present during training. The system must not only identify known identities but also reject novel ones. This setting is used in OpenAnimals [13] and video-based track discovery systems [23], [32].

Single-Shot vs. Multi-Shot: Single-shot evaluation uses only one image per identity in the gallery, while multi-shot provides multiple references. Datasets such as SealID [8] and WildlifeReID-10k [12] support both settings.

Video-Based Evaluation: Video Re-ID requires tracking individuals over sequences. Models are evaluated based on their ability to maintain identity consistency across frames. Methods by Hansen et al. [32] and Jensen et al. [23] apply this protocol to animal movement tracking and unsupervised identity discovery.

C. Challenges in Evaluation

Despite the increasing adoption of standard metrics, several challenges remain in evaluating animal Re-ID models:

- **Lack of Unified Protocols:** Datasets often come with inconsistent splits and metric reporting, making cross-paper comparison difficult.
- **Limited Open-Set Benchmarks:** While open-world Re-ID is critical in practice, few datasets include protocols for unknown identity rejection or novelty detection.
- **Metadata Exclusion:** Contextual factors like GPS, time, and weather are rarely integrated into evaluation, even when models utilize them [24], [20].
- **Cross-Dataset Generalization:** Few studies report performance across multiple datasets or species, limiting insights into model robustness.

D. Toward Better Benchmarking Practices

To advance evaluation in the animal Re-ID community, we propose:

- **Standardized Train/Test Splits:** New datasets should provide fixed protocols for reproducibility.
- **Species-Agnostic Testing:** Cross-species evaluations can assess generalization across morphology and appearance variation [16], [14].

- **Multimodal Evaluation:** When metadata is available, models should be evaluated both with and without it to quantify contextual gain.
- **Open-World Protocols:** Benchmarks should simulate discovery and rejection of novel individuals.
- **Temporal Robustness:** Longitudinal datasets (e.g., SealID, SeaTurtleID2022) should evaluate identity consistency over time.

VIII. Open Challenges and Future Directions

While significant progress has been made in animal re-identification (Re-ID), the field continues to face substantial challenges that limit its real-world applicability, generalization, and scalability. In this section, we highlight key open problems and suggest promising directions for future research.

A. Species Generalization

Most current models are developed and evaluated on a single species or narrow species set. However, practical wildlife monitoring requires systems that can generalize across diverse animal taxa.

Li et al. [17] and Xie et al. [13] explore transformer-based and community-curated approaches to address species diversity, yet cross-species generalization remains limited. Frühner and Tapken [22] demonstrate that transferring person Re-ID methods to multispecies domains is possible, but performance drops significantly without domain-specific fine-tuning.

Future research should emphasize species-agnostic representation learning, using multi-task learning or domain adaptation techniques to build more universal Re-ID systems.

B. Limited Annotations and Label Scarcity

The collection of high-quality labeled data remains a major bottleneck. Manual annotation is labor-intensive, especially in remote, open-world scenarios.

To address this, recent works have adopted few-shot [1], [20], unsupervised [25], and self-supervised [21] methods. Clustering-based pseudo-labeling and active learning pipelines are also promising. Simultaneously, synthetic data generation via GANs [37], [38] may provide cost-effective data augmentation, especially for underrepresented species or rare poses.

C. Environmental Robustness

Real-world wildlife monitoring is conducted in highly variable conditions — different weather, lighting, camera quality, occlusions, and seasonal appearance changes.

Pattern-aware methods like pelage unwrapping [25] or geometric aggregation [2] have improved performance under variation, but significant challenges remain. Transformer-based methods [16], [17] offer some robustness through global context modeling but still require extensive domain-specific training.

Future systems must be resilient to environmental perturbations and robust under incomplete visual evidence. Data augmentation, adversarial training, and domain-invariant representation learning are potential pathways.

D. Benchmark Fragmentation

The lack of standardized datasets, evaluation splits, and reporting protocols limits comparability across studies. Each paper often introduces its own dataset and custom metrics, fragmenting the field.

Standardizing open-world evaluation, as explored in OpenAnimals [13] and SeaTurtleID2022 [7], and releasing metadata-rich datasets [24], [20] with fixed splits can help mitigate this issue. Cross-dataset generalization reporting should become standard practice.

E. Multimodal and Context-Aware Re-ID

Current models focus largely on visual features, ignoring valuable auxiliary cues such as time, GPS, temperature, or even acoustic signals. Choudhury et al. [20] demonstrate that integrating metadata enhances Re-ID performance, especially in open-world and low-data settings.

The field should move toward multimodal Re-ID — fusing vision with environmental, behavioral, or acoustic data. Vision-language models like CLIP [11] also open up opportunities for semantic-level similarity learning and zero-shot retrieval.

F. Scalability and Continual Learning

Many deployed systems must re-identify hundreds or thousands of animals over months or years. This requires:

- **Scalable retrieval**: Efficient index structures and memory-based models
- **Continual learning**: Updating identities without catastrophic forgetting
- **Lifelong learning**: Adapting to new species, appearance changes, and modalities

Incremental metric learning, self-supervised updates, and adaptive memory networks are promising for long-term deployments.

G. Ethical and Ecological Considerations

As Re-ID systems are increasingly used in conservation, ethical deployment must be considered. These include:

- Minimizing disturbance to wildlife (e.g., non-invasive camera placement)
- Ensuring data privacy and ecological transparency
- Preventing misuse in wildlife tracking or poaching

Datasets should include usage guidelines, and model development should align with conservation goals and environmental sustainability.

IX. Conclusion and Outlook

Animal re-identification (Re-ID) has emerged as a vital task in computer vision, enabling long-term monitoring of individual animals for ecological research, conservation, and behavioral analysis. Over the past five years, significant advancements have been made in dataset creation, feature representation learning, similarity modeling, and open-world generalization. However, major challenges remain that limit the scalability and robustness of current systems in real-world deployments.

This survey has provided a comprehensive overview of the state of the art in animal Re-ID from 2020 to 2025. We have reviewed key methods in both closed-world and open-world settings, highlighted the evolution of CNN-based and transformer-based architectures [17], [9], [10], and discussed hybrid and self-supervised learning strategies [25], [21], [20]. We also examined core benchmarks such as WildlifeReID-10k [12], SeaTurtleID2022 [7], and OpenAnimals [13], and summarized current evaluation metrics and protocols [23], [8], [32].

A. Future Research Directions

Several themes are expected to shape the future of this field:

- **Species-Agnostic and Cross-Domain Re-ID:** Models must generalize to unseen species without retraining. Cross-species benchmarks and domain adaptation techniques are crucial [14], [22].
- **Few-Shot and Unsupervised Learning:** Leveraging meta-learning and self-supervised approaches will reduce annotation costs and improve scalability [1], [20], [25].
- **Multimodal Integration:** The fusion of visual features with auxiliary data (e.g., GPS, time, text, audio) will enable more robust and context-aware systems [24], [20], [11].
- **Open-World and Continual Learning:** Systems should support dynamic identity sets, detect novel individuals, and learn incrementally [13], [31], [6].
- **Longitudinal and Temporal Modeling:** Models should account for appearance changes due to growth, aging, or seasonal effects, particularly in conservation settings [7], [8].
- **Ethical AI in Conservation:** Responsible use of Re-ID technology is essential. Collaborations between AI researchers and ecologists must ensure that these tools aid — rather than hinder — wildlife preservation.

B. Outlook

With the increasing availability of high-resolution camera traps, drone imagery, and community-driven datasets, animal Re-ID is poised for rapid growth. By incorporating principles from person Re-ID, few-shot learning, multimodal AI, and ecological ethics, the next generation of animal Re-ID systems can become robust, generalizable, and impactful.

We envision a future where automated Re-ID becomes a standard tool for biodiversity monitoring — enabling non-invasive, real-time, and large-scale tracking of wildlife populations across the globe.

REFERENCES

[1] A. Alahi and et al., "An open-source general purpose machine learning framework for individual animal re-identification using few-shot learning," *Nature Communications*, 2024.

[2] C. Li and et al., "Species-agnostic patterned animal re-identification by aggregating deep local features," *CVPR*, 2024.

[3] L. Zhao and et al., "Zebraid: A patterned animal benchmark for re-identification," *Pattern Recognition*, 2021.

[4] T. Bhattacharya and et al., "Similarity learning networks for animal individual re-identification: An ecological perspective," *Ecological Informatics*, 2022.

[5] M. Feng and et al., "Adaptive feature fusion network for individual amur tiger identification," *Sensors*, 2021.

[6] R. Verma and et al., "Knowledge distillation in siamese networks for wildlife re-identification," *CVPR*, 2023.

[7] L. Qian and et al., "Seaturtleid2022: A long-span dataset for reliable sea turtle re-identification," *Scientific Reports*, 2022.

[8] J. Koskela and et al., "Sealid: Saimaa ringed seal re-identification dataset and methods," *Nature Communications*, 2022.

[9] K. He and et al., "Deep residual learning for image recognition," *CVPR*, 2016.

[10] A. Dosovitskiy and et al., "An image is worth 16x16 words: Transformers for image recognition at scale," *ICLR*, 2021.

[11] A. Radford and et al., "Learning transferable visual models from natural language supervision," *ICML*, 2021.

[12] Y. Xie and et al., "Multispecies animal re-id using a large community-curated dataset," *IJCV*, 2024.

[13] ——, "Openanimals: Revisiting person re-identification for animals towards better generalization," *CVPR*, 2025.

[14] Y. Jiao and et al., "Toward re-identifying any animal: Open-world learning with a generalizable vision transformer," *TPAMI*, 2024.

[15] S. Woo and et al., "Cbam: Convolutional block attention module," *ECCV*, 2018.

[16] F. Xu and et al., "Hierarchical frequency fusion transformer for animal re-identification," *CVPR*, 2023.

[17] S. Li and et al., "Adaptive high-frequency transformer for diverse wildlife re-identification," *NeurIPS*, 2023.

[18] F. Schroff and et al., "Facenet: A unified embedding for face recognition and clustering," *CVPR*, 2015.

[19] A. Hermans and et al., "In defense of the triplet loss for person re-identification," *arXiv preprint arXiv:1703.07737*, 2017.

[20] A. Choudhury and et al., "Meta-feature adapter for few-shot animal re-identification with context awareness," *NeurIPS*, 2023.

[21] X. Zhao and et al., "Indivaid: Learning disentangled representations for unsupervised animal re-identification," *ECCV*, 2023.

[22] J. Frühner and J. Tapken, "From persons to animals: Transferring person re-identification methods to a multi-species animal domain," *CSRN*, 2024.

[23] T. Jensen and et al., "Track-level clustering for unsupervised animal re-identification in video," *CVPR*, 2021.

[24] S. Rao and et al., "Wildlifedatasets: An open-source toolkit for animal re-identification," *arXiv preprint arXiv:2305.10011*, 2023.

[25] M. Andersson and et al., "Unsupervised pelage pattern unwrapping for pose-invariant animal re-identification," *ECCV*, 2022.

[26] J. Chang and et al., "Serial multi-scale fusion for individual identification of amur tigers," *IEEE Transactions on Image Processing*, 2022.

[27] J. Kim and et al., "Vision transformer for lion re-identification," *IJCV*, 2024.

[28] X. Yu and et al., "Landmark-guided feature embedding for animal re-identification," *WACV*, 2022.

[29] Y. Zhang and et al., "Yakre-id-103: A dataset for individual yak re-identification in farm environments," *IEEE Access*, 2022.

[30] T. Wang and et al., "Unsupervised part-based feature alignment for cross-species animal re-identification," *WACV*, 2024.

[31] R. Liu and et al., "Locally aware cross-attention transformers for generalized animal re-id," *ICCV*, 2023.

[32] K. Hansen and et al., "Video-based open-world re-identification of animals in the wild," *CVPR*, 2024.

[33] J. Hu and et al., "Squeeze-and-excitation networks," *CVPR*, 2018.

[34] R. Singh and et al., "Animal re-identification algorithm for posture diversity," *ICCV*, 2023.

[35] W. Liu and et al., "Sphereface: Deep hypersphere embedding for face recognition," *CVPR*, 2017.

[36] H. Wang and et al., "Cosface: Large margin cosine loss for deep face recognition," *CVPR*, 2018.

[37] T. Karras and et al., "Progressive growing of gans for improved quality, stability, and variation," *ICLR*, 2018.

[38] A. Brock and et al., "Large scale gan training for high fidelity natural image synthesis," *ICLR*, 2019.