

A Species-Wise Clustering Pipeline with Salamander Edge Reliability Learning for AnimalCLEF 2026

Fuchuan Ye¹, Xingrun Wang¹, Kaiyi Li¹ and Zhongyuan Han^{1,*}

¹Foshan University

Abstract

This working-notes paper describes our participant submission to AnimalCLEF 2026 for open-set animal individual discovery and re-identification. The system used a separate clustering component for each dataset: the official exact-baseline output for SeaTurtleID2022, a leaderboard-selected fixed Lynx clustering output from development ablations, an unsupervised MegaDescriptor/ConvNeXt ReID fusion for TexasHornedLizards, and a newly trained edge-reliability model for SalamanderID2025. In the Salamander component, a species-specific MLP projector and RandomForest edge classifier combine global similarities, ORB local texture scores, pairwise interactions, and nearest-neighbor rank cues; a three-seed ensemble then accepts high-vote Union-Find merges under a maximum cluster-size constraint. The final submitted run obtained a Public ARI of **0.26227** and a final Private ARI of **0.28905**. The largest public-leaderboard gain after the Texas stage came from replacing fixed-threshold Salamander merging with supervised edge reliability learning and multi-seed voting.

Keywords

AnimalCLEF 2026, LifeCLEF, animal re-identification, open-set discovery, clustering, edge classification, Adjusted Rand Index

1. Introduction

Identifying individual animals from images is a long-standing requirement in ecological monitoring, population estimation, behavioral analysis, and conservation management. Traditional individual identification often depends on manual inspection of natural markings, tags, scars, or photographic catalogues. This process is accurate when performed by experts, but it is difficult to scale to the volume of images produced by camera traps, drones, citizen-science platforms, and long-term field surveys. Computer vision methods for animal re-identification aim to reduce this bottleneck by matching images of the same individual across time, viewpoint, illumination, device, and background changes.

Animal re-identification differs from ordinary species classification. The target label is not the species name, but the identity of a particular animal within a species. This makes the task closer to instance-level retrieval and open-set recognition: a test image may correspond to an already known individual, but it may also show a new individual that has not appeared in the reference set. Public benchmarks and toolkits such as SeaTurtleID2022 [1], CzechLynx [2], WildlifeDatasets [3], and WildlifeReID-10k [4] have helped standardize data access and evaluation for this setting. The previous AnimalCLEF edition evaluated individual recognition for lynxes, salamanders, and sea turtles [5]. LifeCLEF and AnimalCLEF provide a complementary benchmark context by evaluating complete participant systems under an official leaderboard protocol [6, 7].

AnimalCLEF 2026 focuses on discovery and re-identification of individual animals from multiple species. Participants are given image metadata and must output a `cluster` label for every test image. Images predicted to belong to the same animal should share a cluster label; images of different animals should be assigned different labels. The official metric is the Adjusted Rand Index (ARI), which compares the pairwise consistency of the predicted partition with the hidden ground-truth partition [8]. This metric penalizes both splitting one true individual into multiple clusters and merging different individuals

CLEF 2026 Working Notes

*Corresponding author.

✉ yfc2097639788@gmail.com (F. Ye); wangxingrun@fosu.edu.cn (X. Wang); liky8502@gmail.com (K. Li);

hanzhongyuan@gmail.com (Z. Han)

ORCID 0009-0002-1467-0358 (F. Ye); 0000-0002-7364-198X (X. Wang); 0009-0006-2500-1036 (K. Li); 0000-0001-8960-9872 (Z. Han)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

into one predicted cluster. A strict global similarity threshold can fragment one individual under pose or illumination changes, while a loose threshold can cause transitive over-merging. Therefore, the central modeling problem is not simply choosing a number of clusters, but deciding which image pairs are reliable same-individual links.

The final system therefore uses species-wise components. SeaTurtle and Lynx use fixed clustering outputs selected during leaderboard development, Texas uses unsupervised feature fusion, and Salamander receives the trainable component: supervised edge reliability learning followed by conservative graph merging.

The contributions of this working-notes paper are:

- We describe the submitted system, including which outputs were reused and which components were trained or modified.
- We present a Salamander-specific edge reliability module for selecting conservative same-individual links before graph merging.
- We report leaderboard ablations and negative results that show where each change helped.

2. Related Work

2.1. Animal Re-Identification Benchmarks

Recent animal re-identification benchmarks cover both single-species case studies and multi-species resources. Single-species datasets are often designed around stable visual cues: facial and scute patterns for sea turtles in SeaTurtleID2022 [1], coat patterns and camera-trap metadata for Eurasian lynx in CzechLynx [2], and ventral spot patterns for Texas horned lizards [9]. These datasets show that the useful identity signal can be species dependent and may appear at different spatial scales, from global body shape to small local markings.

Multi-species resources address the complementary problem of training and evaluating models that transfer across taxa. WildlifeDatasets standardizes access to many animal re-identification datasets and provides tooling for feature extraction and evaluation [3], while WildlifeReID-10k collects more than ten thousand individual animals across many species for representation learning and benchmark evaluation [4]. AnimalCLEF builds on this ecosystem by evaluating complete participant systems under a challenge protocol with multiple species and open-set behavior [6, 7]. In this setting, the submitted output is a partition of the test images rather than only a closed-set identity classification.

2.2. Global Embeddings and Metric Learning

Deep metric learning is widely used in re-identification because it learns an embedding space in which same-identity samples are close and different-identity samples are separated. Triplet loss is one common formulation for this objective [10]. In animal ReID, MegaDescriptor, provided through the WildlifeDatasets/WildlifeTools ecosystem, is a strong general-purpose feature extractor for individual animal recognition [3]. WildFusion further studies calibrated fusion of deep and local similarity scores for individual animal identification [11]. Our system uses MegaDescriptor as the main global feature and also uses an auxiliary *ConvNeXt ReID embedding*: a ConvNeXt-Base representation [12] trained on WildlifeReID-10k for animal individual re-identification. We use this descriptive term throughout the remainder of the paper instead of the vaguer phrase “general ReID embedding”.

2.3. Local Matching and Graph-Based Clustering

For animals whose identities are expressed by spots, stripes, scars, or fine local patterns, global image embeddings can miss discriminative details. Local descriptors such as ORB [13] offer a lightweight way to compare local texture structures, while SIFT and RANSAC provide classical descriptor and geometric-verification tools [14, 15]. More recent learned matchers, including LightGlue-style pipelines [16], can be stronger in rigid or near-rigid matching, but our experiments showed that they did not improve

the Salamander leaderboard score in this task. Once reliable image-pair scores are available, graph or Union-Find merging provides a natural way to form clusters. However, one false-positive edge can propagate through transitive closure, so our final Salamander module adds vote and cluster-size constraints to reduce risky merges.

3. Task and Data

3.1. Task Setting and Dataset Labels

AnimalCLEF 2026 is a discovery and re-identification task for clustering individual animals from images [7]. The submission used the Kaggle release and metadata available to participants. The metadata used by our system contained four dataset labels:

- **LynxID2025**: Eurasian lynx (*Lynx lynx*), linked to the CzechLynx individual-identification data source [2];
- **SalamanderID2025**: fire salamander (*Salamandra salamandra*), introduced for the AnimalCLEF benchmark setting in the previous challenge edition [5];
- **SeaTurtleID2022**: loggerhead sea turtle (*Caretta caretta*) [1];
- **TexasHornedLizards**: Texas horned lizard (*Phrynosoma cornutum*), motivated by the published study of individual recognition from ventral spot patterns [9].

LynxID2025, SalamanderID2025, and SeaTurtleID2022 correspond to species used in the previous AnimalCLEF edition [5], while TexasHornedLizards is the newly introduced dataset in the 2026 release used by our submission. For LynxID2025, SalamanderID2025, and SeaTurtleID2022, the training/database split contains identity-labeled images. The test split contains a mixture of known and previously unseen individuals; therefore, a system must support both lookup and discovery. TexasHornedLizards appears as a test-only species in the release we used, so we treat it as an unsupervised discovery subtask. The metadata provide image identifiers and paths, dataset/species labels, split information, and identity labels for training images where available. We clustered images within each dataset label, matching the submitted CSV format and preventing cross-species links.

3.2. Input, Output, and Metric

The input to the submitted system consists of test images and metadata. The required output is a CSV file with one row per test image and two columns:

```
image_id, cluster
```

The `image_id` values must follow the official test/sample-submission order, and `cluster` is an arbitrary predicted identity-group label for the corresponding test image. ARI is computed by comparing the submitted partition with the hidden ground-truth partition. Because ARI measures pairwise partition agreement, our Salamander module learns reliable pairwise edges before graph merging.

4. Submitted System

4.1. Overview and Component Sources

Table 1 summarizes the final system and the source of each component. The system treats each species as an open-set clustering problem: labeled reference identities are used only where they help learn representations or edge reliability, and the final output is always a partition of the test images. SeaTurtle and Lynx use reused fixed outputs, Texas uses an author-modified unsupervised fusion, and Salamander uses the newly trained edge-learning component.

Table 1

Final submitted system by species.

Species	Final strategy	Component source	Key parameters	New
SeaTurtleID2022	Official exact baseline output	Reused official baseline output	MegaDescriptor-L-384, MiewID branch, WildFusion-style fusion, HDBSCAN in the baseline configuration	No
LynxID2025	Fixed Lynx clustering output kept from development ablations	Reused leaderboard-selected output	946 images, 226 clusters, 129 singletons; Public ARI 0.21928 in single-species swap against the exact baseline	No
TexasHornedLizards	Unsupervised similarity fusion	Author-modified component	$0.85S^{Mega} + 0.15S^{ConvNeXt}$, distance threshold 0.27	Partial
SalamanderID2025	Edge reliability learning and constrained merging	Newly trained component	MLP projector, ORB score, RandomForest, three seeds, two-vote edge rule, 36 accepted merges, $K_{\max} = 4$	Yes

4.2. Official Exact Baseline and Lynx Selection Rationale

The official exact baseline referred to in this paper is the baseline output reproduced from the competition materials. It combined MegaDescriptor-L-384, a MiewID feature branch, WildFusion-style similarity fusion [11], and HDBSCAN clustering [17]. We reused its SeaTurtleID2022 labels without retraining because our supervised-projector replacements fragmented the public-leaderboard partition into more clusters and singletons.

For LynxID2025, we did not train a new Lynx model for the final run. Instead, the final submission reused the Lynx labels from an earlier leaderboard-selected clustering output. To verify the effect of this reused component, we ran a single-species swap test in which only the Lynx labels of the official exact-baseline submission were replaced. This swap covered 946 Lynx test images and produced 226 predicted clusters with 129 singletons; it raised the Public ARI of that diagnostic submission from 0.20330 to 0.21928. A later MegaDescriptor-only agglomerative-clustering diagnostic reproduced the same 226-cluster/129-singleton structure at distance threshold 0.38, suggesting that the selected Lynx partition was mainly a conservative global-embedding clustering rather than a new trained method. By comparison, a Lynx-specific projector candidate scored 0.25904 Public ARI in the same multi-species system context, below the 0.26067 run that kept the earlier Lynx clustering before adding the final Salamander ensemble. These comparisons led us to keep Lynx fixed and reserve the trained projector for Salamander experiments.

4.3. TexasHornedLizards Similarity Fusion

TexasHornedLizards had no identity-labeled training subset available in the release we used. We used a simple unsupervised fusion of two global similarities. Let S_{ij}^{Mega} be the cosine similarity between MegaDescriptor features and $S_{ij}^{ConvNeXt}$ be the cosine similarity between the WildlifeReID-10k-trained ConvNeXt ReID embeddings. We used:

$$S_{ij}^{Texas} = 0.85S_{ij}^{Mega} + 0.15S_{ij}^{ConvNeXt}. \quad (1)$$

The distance matrix was $1 - S_{ij}^{Texas}$, and the submitted implementation applied agglomerative clustering with average linkage and a distance threshold of 0.27. Table 3 includes this step as the Texas fusion stage.

4.4. Salamander-Specific Projection

The Salamander module starts from the concatenation of MegaDescriptor and ConvNeXt ReID features. This concatenation is intended to align the strong general identity signal of MegaDescriptor with an auxiliary representation learned from a large multi-species ReID dataset. A two-layer MLP maps the concatenated vector into a 512-dimensional salamander-specific embedding. Batch normalization follows each linear layer: the hidden-layer BatchNorm is applied before ReLU and dropout, and the output-layer BatchNorm is applied before L2 normalization:

```

concat([Mega feature, ConvNeXt ReID feature])
-> Linear(in_dim, 1024)
-> BatchNorm1d(1024)
-> ReLU
-> Dropout(0.20)
-> Linear(1024, 512)
-> BatchNorm1d(512)
-> L2 Normalization

```

The classifier head is used only during training. The training loss is:

$$\mathcal{L} = \mathcal{L}_{CE} + 0.5\mathcal{L}_{Triplet}, \quad (2)$$

where the triplet margin is 0.25. The final ensemble used embedding dimension 512, hidden dimension 1024, dropout 0.20, AdamW with learning rate 10^{-3} , batch sampling with `batch_p=32`, `batch_k=2`, and training seeds 42, 2026, and 3407.

4.5. Salamander Edge Reliability Learning

Candidate Salamander edges are generated from the union of the top- k nearest neighbors under MegaDescriptor similarity and projected-embedding similarity, with $k = 40$ in the final ensemble. Each unordered image pair is represented once before feature construction. During training, candidate pairs from the Salamander training identities are labeled positive when the two images have the same identity and negative otherwise. At test time the same feature construction is applied to unlabeled candidate pairs. Absolute similarity scores can be sensitive to blur, illumination, cropping, and occlusion; therefore, we add nearest-neighbor rank statistics as relative reliability cues rather than relying only on raw cosine values. For every candidate pair (i, j) , we compute:

- MegaDescriptor cosine similarity;
- salamander-projector cosine similarity;
- ORB local texture score;
- pairwise interaction terms, including Mega times local, projector times local, and Mega times projector;
- nearest-neighbor rank statistics under both similarity matrices;
- mutual top-5 and mutual top-10 indicators.

The ORB local texture score is:

$$s_{local}(i, j) = 0.5 \cdot \frac{N_{good}}{\min(N_i, N_j)} + 0.5 \cdot \min\left(\frac{N_{good}}{50}, 1\right), \quad (3)$$

where N_{good} is the number of ratio-test matches and N_i, N_j are the descriptor counts of the two images. This normalized score rewards pairs with many reliable local matches while reducing the effect of different descriptor counts across images. The edge classifier is a RandomForest [18] trained on same-identity and different-identity candidate pairs with `max_depth=7`, `min_samples_leaf=6`, and `class_weight=balanced_subsample`. The negative class consists of candidate pairs whose two training images have different identities, and the RandomForest class weighting handles the remaining class imbalance. The single-model version used 600 trees; the final ensemble used 700 trees per seed.

4.6. Topological-Constrained Ensemble Merging

Single false-positive edges can trigger large transitive errors when graph components are merged. To reduce this risk, we combine model agreement with a hard topological bound on the resulting cluster

Table 2

Final submission metadata.

Item	Value
Team name / Kaggle ID	ChaBuDuoXianSheng (Kaggle display name: chabuduoXiansheng)
Public score	0.26227 ARI (public leaderboard split; 48% of the test set)
Private score	0.28905 ARI (hidden private leaderboard split; final ranking phase)
Metric	Adjusted Rand Index
Leaderboard phase	Public test-phase score for model selection; private final-leaderboard score for official rank
Submitted artifact	Final Kaggle CSV generated from the species-wise pipeline; internal filenames are omitted for reproducibility clarity
Official rank	63
Code repository	https://github.com/SoBigLikeMe/animalclef2026-code

size. For each seed m , the classifier predicts a same-individual probability $p_{ij}^{(m)}$ for candidate edge (i, j) . We aggregate probabilities over the models that produced the edge:

$$\bar{p}_{ij} = \frac{1}{M_{ij}} \sum_{m=1}^{M_{ij}} p_{ij}^{(m)}, \quad (4)$$

where M_{ij} is the number of seed models that include this edge. A seed that did not produce a candidate pair contributes no probability and no vote for that pair. The final system keeps only edges with at least two votes:

$$e_{ij} \in \mathcal{E}_{final} \quad \text{iff} \quad M_{ij} \geq 2 \quad (5)$$

Edges are sorted by the ensemble score and merged with Union-Find. A merge is rejected if it would create a cluster larger than $K_{\max} = 4$. The parameter `target_merge` denotes the maximum number of accepted Union-Find merge operations; after that many successful merges, the merging process stops. The final accepted setting was:

```
min_votes = 2
target_merge = 36
K_max = 4
```

This setting produced 470 Salamander clusters and 356 Salamander singletons in the final submission. Allowing one additional accepted Salamander merge reduced the Public ARI from 0.26227 to 0.26202, so we kept the more conservative 36-merge setting.

5. Experimental Setup

5.1. Evaluation Protocol and Submission Metadata

All scores reported in this paper are AnimalCLEF 2026 leaderboard ARI values, not local validation scores unless explicitly stated otherwise. During the test phase, Kaggle displayed the public score computed on 48% of the test set. The final ranking used the hidden private leaderboard split. Table 2 summarizes the final-run metadata and separates the public and private leaderboard phases.

5.2. Implementation and Selection Details

The main input features are MegaDescriptor and the WildlifeReID-10k-trained ConvNeXt ReID embedding. They are used directly for TexasHornedLizards similarity fusion and concatenated before the Salamander projector. All similarity computation and clustering are performed within the same species/sub-dataset. For supervised Salamander components, the projector uses embedding dimension 512, hidden dimension 1024, dropout 0.20, AdamW with learning rate 10^{-3} , `batch_p=32`, `batch_k=2`,

Table 3

Step-by-step public-leaderboard ablation results from the AnimalCLEF 2026 development trajectory. All rows use ARI on the public leaderboard split; none are local-validation scores.

Stage	Changed component	Metric	Public score	Split/phase
Initial solution	Unified early MegaDescriptor clustering	ARI	0.15448	Public LB split
Official exact baseline	MegaDescriptor-L-384 + MiewID + WildFusion-style fusion + HDBSCAN	ARI	0.20330	Public LB split
Species replacement	Per-species output swap for Lynx/Texas	ARI	0.23121	Public LB split
Texas fusion	Texas: 0.85 Mega + 0.15 ConvNeXt ReID, threshold 0.27	ARI	0.25097	Public LB split
Salamander ORB q98	Fixed-threshold ORB local texture merge	ARI	0.25424	Public LB split
Salamander ORB q983/q984	Refined fixed-threshold local merge	ARI	0.25507	Public LB split
SIFT/RANSAC	Local geometric verification replacing ORB logic	ARI	0.25327	Public LB split
ALIKED/LightGlue	Learned local matching variant	ARI	0.24964	Public LB split
ORB edge classifier	Mega + ORB + rank features, RandomForest with 35 accepted merges	ARI	0.25771	Public LB split
Projector edge classifier	Add Salamander-specific projector features	ARI	0.26067	Public LB split
Final ensemble	Three seeds, two-vote edge rule, 36 accepted merges, $K_{\max} = 4$	ARI	0.26227	Public LB split

and triplet margin 0.25. The final ensemble uses seeds 42, 2026, and 3407. We selected variants using public leaderboard ARI together with cluster and singleton counts. Reused components are limited to the official SeaTurtle baseline labels and the fixed Lynx labels described in Table 1; the Texas fusion weights and the Salamander edge-learning pipeline were implemented by the authors for this submission.

5.3. Baselines and Compared Variants

We use the following baselines and variants:

- **Early MegaDescriptor clustering:** an initial unified clustering system using MegaDescriptor features and one global clustering rule. It is included only as an early score reference.
- **Official exact baseline:** the exact baseline reproduced from the competition materials, using MegaDescriptor-L-384, MiewID, WildFusion-style fusion, and HDBSCAN. Its outputs are used without retraining.
- **Species-wise replacement:** a system that keeps the exact baseline for some species and swaps in higher-scoring per-species outputs for others. The Lynx row uses a fixed clustering output from an earlier submission.
- **Texas fusion:** the weighted MegaDescriptor/ConvNeXt ReID fusion described above.
- **Salamander ORB threshold:** fixed-threshold local texture merging.
- **Salamander edge classifier:** supervised RandomForest edge prediction using MegaDescriptor, ORB, and rank features.
- **Salamander projector edge classifier:** the same edge classifier augmented with salamander-specific projected embeddings.
- **Final ensemble:** multi-seed projector/classifier voting with constrained Union-Find.

6. Results and Ablations

6.1. Ablation Results

Table 3 summarizes the score trajectory. The rows are not independent controlled trials on a hidden validation set; they are the documented leaderboard ablations used to select the final working-notes submission.

The score trajectory shows where the final improvement came from. After the Texas fusion row, the remaining gains are tied to Salamander changes: ORB thresholding improves over the Texas stage, and the RandomForest edge classifier, projector features, and three-seed voting give the best public score.

Table 4

Salamander-side public-leaderboard gain decomposition after the Texas fusion stage. The metric is ARI on the public leaderboard split.

Retained stage	Public ARI	Gain	Main added evidence
Texas fusion baseline	0.25097	–	No Salamander-specific edge learning
Fixed ORB local merge	0.25507	+0.00410	Local texture matches select additional Salamander links
RandomForest edge classifier	0.25771	+0.00264	Similarity, ORB, interaction, and rank features
Projector edge classifier	0.26067	+0.00296	Salamander-specific projected embedding
Final three-seed ensemble	0.26227	+0.00160	Seed voting plus constrained Union-Find merging

Table 4 isolates these Salamander-side changes and reports the public-leaderboard gain relative to the preceding retained stage. The lower Lynx and SeaTurtle projector results in Section 6.2 explain why those components were left unchanged.

6.2. Negative Results

Vision foundation features. DINO-family foundation-feature variants [19] scored around 0.248 during development, below the Texas-fusion and Salamander-edge systems, so they were left out of the final submission.

Learned or geometric local matchers. SIFT [14] plus RANSAC [15] obtained 0.25327, and ALIKED [20] plus LightGlue [16] obtained 0.24964. Both were below the highest-scoring ORB-based fixed-threshold score and well below the edge-classifier variants. For this dataset, the non-rigid deformation, body bending, local occlusion, and viewpoint changes in salamander images appear to make these matchers less effective than the ORB-based features used by the final edge classifier.

Projector transfer to other species. A Lynx projector variant using a 70/30 MegaDescriptor/projector fusion scored 0.25904, lower than the corresponding system that kept the earlier Lynx clustering. SeaTurtle projector variants produced many more clusters and singletons than the official exact baseline, indicating fragmentation. In the final submission, the projector is therefore used only for SalamanderID2025.

7. Discussion and Limitations

The final system is a competition pipeline built around one trained component. The species-wise selection kept fixed outputs when replacement experiments reduced the score, used unsupervised fusion for the test-only lizard subset, and concentrated supervised modeling on SalamanderID2025. This design follows the ARI objective: instead of relying on one global threshold, the Salamander module estimates pairwise edge reliability and then applies conservative graph merging to prevent one false edge from producing a large transitive error.

The approach has limitations. The Salamander edge classifier is trained on candidate pairs from the training identities and may overfit the candidate-edge distribution; multi-seed voting and cluster-size constraints reduce but do not eliminate this risk. Variant selection relied on public leaderboard feedback because the ground-truth test identities were hidden, so Table 3 is best read as a development trajectory rather than an independent validation study. The Lynx component is a fixed clustering output chosen for lower fragmentation relative to the official exact baseline. External resources used by the system include MegaDescriptor/WildlifeDatasets features and a WildlifeReID-10k-trained ConvNeXt-Base ReID embedding.

8. Conclusion

We presented our AnimalCLEF 2026 submission. The final system combines fixed SeaTurtle and Lynx outputs, Texas feature fusion, and Salamander edge-reliability learning. The Salamander component

learns same-individual edges from projected global embeddings, ORB local texture scores, and nearest-neighbor topology, then merges high-vote edges under a maximum cluster-size constraint. The final run achieved a Public ARI of 0.26227, a Private ARI of 0.28905, and official rank 63. In the public leaderboard trajectory, the best gains after the Texas row came from supervised edge classification, projected features, and multi-seed voting.

Acknowledgments

This work is supported by the Guangdong Province Basic and Applied Basic Research Fund (Guangdong-Foshan Joint Funds, 2024A1515140026).

Declaration on Generative AI

During the preparation of this manuscript version, the author(s) used OpenAI Codex/ChatGPT for translation, editing, and formatting assistance according to the CEUR-WS ceurart template.

References

- [1] L. Adam, V. Cermak, K. Papafitsoros, L. Pícek, SeaTurtleID2022: A long-span dataset for reliable sea turtle re-identification, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2024, pp. 7146–7156.
- [2] L. Pícek, J. Straka, M. Jirik, E. Belotti, M. Dula, J. Krausova, M. Bojda, V. Cermak, L. Bufka, R. Dvorak, L. Hrdy, V. Kocourek, J. Labuda, L. Toman, V. Trulik, M. Vana, M. Kutal, CzechLynx: A dataset for individual identification and pose estimation of the eurasian lynx, Scientific Data 13 (2026) 511. doi:10.1038/s41597-026-06853-9.
- [3] V. Cermak, L. Pícek, L. Adam, K. Papafitsoros, WildlifeDatasets: An open-source toolkit for animal re-identification, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2024, pp. 5953–5963.
- [4] L. Adam, V. Cermak, K. Papafitsoros, L. Pícek, WildlifeReID-10k: Wildlife re-identification dataset with 10k individual animals, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2025, pp. 2124–2134. doi:10.1109/CVPRW67362.2025.00197.
- [5] L. Adam, K. Papafitsoros, R. Kovar, V. Cermak, L. Pícek, Overview of AnimalCLEF 2025: Recognizing individual animals in images, in: CLEF 2025 Working Notes, volume 4038 of *CEUR Workshop Proceedings*, 2025, pp. 2898–2908. URL: https://ceur-ws.org/Vol-4038/paper_231.pdf.
- [6] L. Pícek, L. Adam, S. Kahl, R. Bossy, L. Chrobak, H. Goëau, K. Papafitsoros, H. Klinck, W.-P. Vellinga, R. Planqué, T. Denton, K. Barnard, C. Nédellec, L. Deléger, M. Courtin, G. Martellucci, I. Moummad, F. Vinatier, P. Bonnet, A. Joly, Overview of LifeCLEF 2026: Ai challenges for biodiversity understanding and ecosystem management, in: International Conference of the Cross-Language Evaluation Forum for European Languages (CLEF), Springer, 2026.
- [7] L. Adam, K. Papafitsoros, D. A. Williams, D. Biffi, L. Pícek, Overview of AnimalCLEF 2026: Discovery and re-identification of individual animals, in: Working Notes of CLEF 2026 - Conference and Labs of the Evaluation Forum, 2026.
- [8] L. Hubert, P. Arabie, Comparing partitions, Journal of Classification 2 (1985) 193–218. doi:10.1007/BF01908075.
- [9] D. Biffi, M. R. Tucker, A. Ackel, D. A. Williams, Identification of individual texas horned lizards (phrynosoma cornutum) using genotypes and ventral spot patterns, Ecology and Evolution 15 (2025) e71167. doi:10.1002/ece3.71167.
- [10] A. Hermans, L. Beyer, B. Leibe, In defense of the triplet loss for person re-identification, arXiv preprint arXiv:1703.07737 (2017). doi:10.48550/arXiv.1703.07737.
- [11] V. Cermak, L. Pícek, L. Adam, L. Neumann, J. Matas, WildFusion: Individual animal identification with calibrated similarity fusion, in: Computer Vision – ECCV 2024 Workshops, vol-

ume 15624 of *Lecture Notes in Computer Science*, Springer, Cham, 2025, pp. 18–36. doi:10.1007/978-3-031-92387-6_2.

- [12] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, S. Xie, A convnet for the 2020s, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 11976–11986.
- [13] E. Rublee, V. Rabaud, K. Konolige, G. Bradski, ORB: An efficient alternative to SIFT or SURF, in: *Proceedings of the IEEE International Conference on Computer Vision*, 2011, pp. 2564–2571. doi:10.1109/ICCV.2011.6126544.
- [14] D. G. Lowe, Distinctive image features from scale-invariant keypoints, *International Journal of Computer Vision* 60 (2004) 91–110. doi:10.1023/B:VISI.0000029664.99615.94.
- [15] M. A. Fischler, R. C. Bolles, Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography, *Communications of the ACM* 24 (1981) 381–395. doi:10.1145/358669.358692.
- [16] P. Lindenberger, P.-E. Sarlin, M. Pollefeys, LightGlue: Local feature matching at light speed, in: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 17627–17638.
- [17] L. McInnes, J. Healy, S. Astels, hdbscan: Hierarchical density based clustering, *Journal of Open Source Software* 2 (2017) 205. doi:10.21105/joss.00205.
- [18] L. Breiman, Random forests, *Machine Learning* 45 (2001) 5–32. doi:10.1023/A:1010933404324.
- [19] M. Oquab, T. Darcet, T. Moutakanni, H. V. Vo, M. Szafraniec, V. Khalidov, P. Fernandez, D. Haziza, F. Massa, A. El-Nouby, M. Assran, N. Ballas, W. Galuba, R. Howes, P.-Y. Huang, S.-W. Li, I. Misra, M. Rabbat, V. Sharma, G. Synnaeve, H. Xu, H. Jegou, J. Mairal, P. Labatut, A. Joulin, P. Bojanowski, DINOv2: Learning robust visual features without supervision, *Transactions on Machine Learning Research* (2024).
- [20] X. Zhao, X. Wu, W. Chen, P. C. Y. Chen, Q. Xu, Z. Li, ALIKED: A lighter keypoint and descriptor extraction network via deformable transformation, *IEEE Transactions on Instrumentation and Measurement* 72 (2023) 1–16. doi:10.1109/TIM.2023.3271000.